OM of: US-09-664-641-10 to: SPTREMBL_19:* out_format : pfs

Date: Jun 4, 2002 5:49 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09664641/runat_04062002_110933_9250/app_query.fasta_1.3697
-DB=SPTREMBL_19 -QFMT=fastan -SUFFIX=rspt -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09664641_@CGN1_1_312 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-664-641-10
Query length: 3580
Database: SPTREMBL_19:*
Database sequences: 562222
Database length: 172994929
Search time (sec): 230.300000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
sp_rodent:Q9Z0W6	+	4059.50	5875.12	0.0	1056	! Q9z0w6 mus musculus (mouse). p
sp_human:O15404	+	3841.00	5561.10	0.0	744	! O15404 homo sapiens (human). ca
sp_vertebrate:Q90WJ3	+	3425.50	4951.40	1.5e-268	1256	! Q90wj3 xenopus laevis (african
sp_human:Q96HP2	+	2039.00	2947.67	1.9e-156	391	! Q96hp2 homo sapiens (human). un
sp_invertebrate:Q9VUB6	+	852.00	1205.56	4.6e-60	1798	! Q9vub6 drosophila melanogast
sp_invertebrate:Q94046	+	524.00	734.07	1.4e-33	1074	! Q94046 caenorhabditis elegan
sp_invertebrate:Q9U370	+	524.00	734.05	1.4e-33	1076	! Q9u370 caenorhabditis elegan
sp_human:Q14676	+	342.50	463.12	8.9e-19	2089	! Q14676 homo sapiens (human). k
sp_human:Q96QC2	+	340.50	460.21	1.3e-18	2090	! Q96qc2 homo sapiens (human). k
sp_plant:Q9LVF7	+	316.50	432.68	8.9e-17	1041	! Q9lvf7 arabidopsis thaliana (m
sp_rodent:Q91VX3	+	283.50	382.38	4.5e-14	1296	! Q91vx3 mus musculus (mouse). s
sp_plant:Q9ZR18	+	265.50	361.78	1.1e-12	765	! Q9zr18 arabidopsis thaliana (mc
sp_human:Q9UEB9	+	263.50	352.22	2.0e-12	1435	! Q9ueb9 homo sapiens (human). c
sp_human:Q92547	+	263.50	351.40	2.0e-12	1550	! Q92547 homo sapiens (human). k
sp_fungi:Q9P5J5	+	229.00	307.74	1.0e-09	838	! Q9p5j5 neurospora crassa. conse
sp_invertebrate:Q9VFR7	+	225.50	303.46	1.9e-09	776	! Q9vfr7 drosophila melanogaste
sp_invertebrate:Q24163	+	220.50	291.76	5.5e-09	1180	! Q24163 drosophila melanogast
sp_invertebrate:Q9VED3	+	220.50	291.71	5.6e-09	1186	! Q9ved3 drosophila melanogast
sp_invertebrate:O44011	+	219.50	288.08	7.2e-09	1457	! O44011 dictyostelium discoid
sp_invertebrate:Q9W4E8	+	216.00	293.94	9.6e-09	517	! Q9w4e8 drosophila melanogaste
sp_plant:Q93VQ3	+	215.50	283.70	1.4e-08	1272	! Q93vq3 oryza sativa (rice). hy
sp_invertebrate:O76853	+	214.00	293.23	1.3e-08	420	! O76853 dictyostelium discoide
sp_fungi:P79025	+	214.00	291.75	1.4e-08	483	! P79025 emericella nidulans (asp
sp_virus:Q98148	+	214.00	282.47	1.9e-08	1162	! Q98148 kaposi's sarcoma-associ
sp_invertebrate:Q9VP17	+	213.50	274.60	2.6e-08	2285	! Q9vp17 drosophila melanogast
sp_vertebrate:Q90YI2	+	212.50	275.54	2.9e-08	1822	! Q90yi2 fugu rubripes (japanese
sp_invertebrate:Q9Y0C9	+	212.00	283.02	2.4e-08	838	! Q9y0c9 dictyostelium discoide
sp_invertebrate:O18346	+	210.50	278.87	3.4e-08	1010	! O18346 drosophila hydei (fru
sp_virus:Q9QR71	+	210.00	276.96	3.9e-08	1129	! Q9qr71 kaposi's sarcoma-associ
sp_invertebrate:Q9VMS2	+	208.50	266.64	6.7e-08	2439	! Q9vms2 drosophila melanogast
sp_invertebrate:Q95ZG5	+	207.00	262.29	9.6e-08	2994	! Q95zg5 dictyostelium discoid
sp_invertebrate:Q9V7C5	+	206.50	273.98	6.9e-08	925	! Q9v7c5 drosophila melanogaste
sp_invertebrate:Q95YM8	+	206.50	268.20	8.4e-08	1598	! Q95ym8 apis mellifera (honey
sp_invertebrate:Q17641	-	205.50	288.88	4.8e-08	197	! Q17641 caenorhabditis elegans
sp_invertebrate:Q9VN59	+	205.00	271.70	9.2e-08	934	! Q9vn59 drosophila melanogaste
sp_invertebrate:Q94485	+	204.50	274.64	8.9e-08	660	! Q94485 dictyostelium discoide
sp_invertebrate:Q9N8T2	-	203.00	255.01	2.1e-07	3439	! Q9n8t2 trypanosoma brucei. p
sp_human:O15410	+	202.50	271.86	1.3e-07	652	! O15410 homo sapiens (human). ca
sp_human:O75557	+	202.50	259.89	1.9e-07	2023	! O75557 homo sapiens (human). d
sp_human:Q9UND7	+	202.50	259.89	1.9e-07	2023	! Q9und7 homo sapiens (human). d
sp_virus:Q9DUN0	+	201.00	265.41	2.0e-07	976	! Q9dun0 kaposi's sarcoma-asso
sp_rodent:Q64507	-	200.00	279.24	1.4e-07	230	! Q64507 mus musculus (mouse).
sp_human:Q9BT44	+	200.00	268.27	2.0e-07	649	! Q9bt44 homo sapiens (human).
sp_virus:O40947	+	200.00	262.80	2.5e-07	1089	! O40947 kaposi's sarcoma-ass
sp_invertebrate:Q9U9S7	+	200.00	255.74	3.1e-07	2123	! Q9u9s7 dictyostelium disc

seq_name: sp_rodent:Q9Z0W6

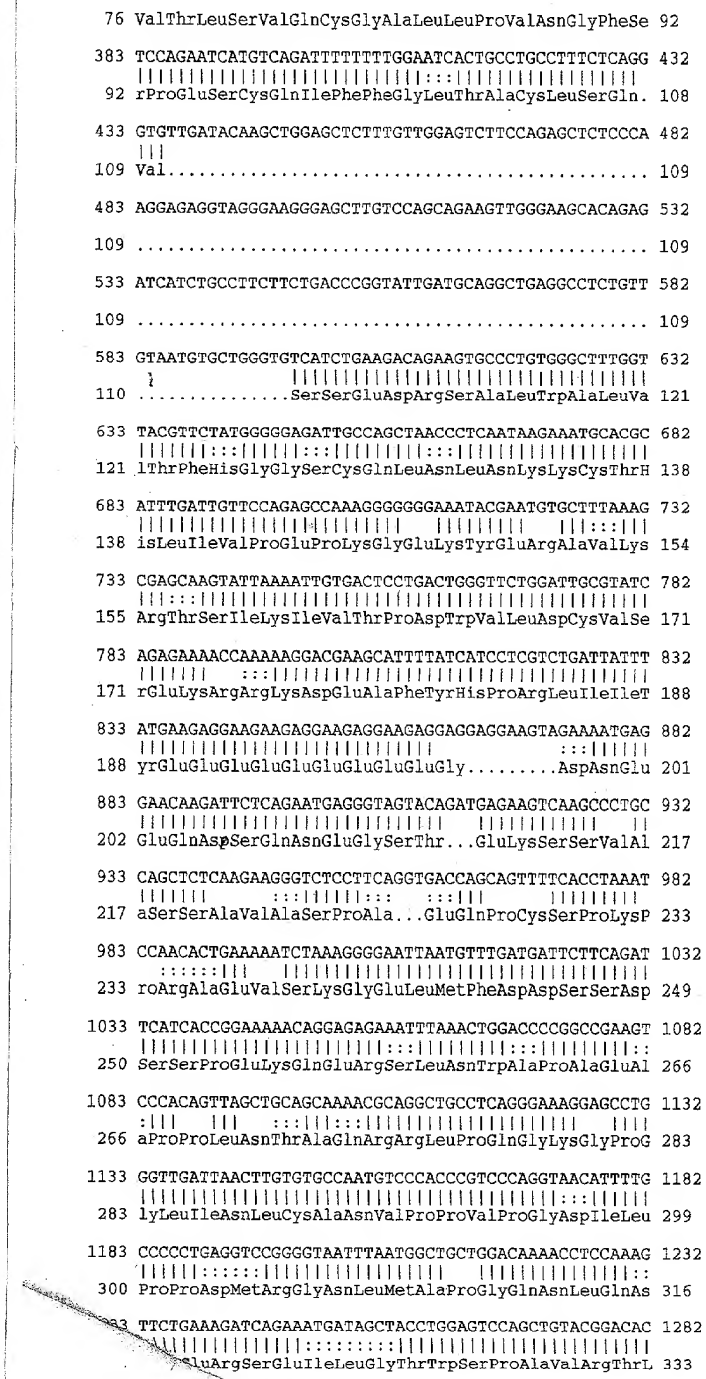
seq_documentation_block:
ID Q9Z0W6 PRELIMINARY; PRT; 1056 AA.
AC Q9Z0W6;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PAX TRANSCRIPTION ACTIVATION DOMAIN INTERACTING PROTEIN PTIP.
GN PAXIP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB;
RX MEDLINE=20368635; PubMed=10908331;
RA Lechner M.S., Levitan I., Dressler G.R.;
RT "PTIP, a novel BRCT domain-containing protein interacts with Pax2 and
RT is associated with active chromatin.";
RL Nucleic Acids Res. 28:2741-2751(2000).
DR EMBL; AF104261; AAD17923.1; -.
DR MGD; MGI:1890430; Paxip1.
DR InterPro; IPR001357; BRCT.
DR Pfam; PF00533; BRCT; 6.
DR SMART; SM00292; BRCT; 5.
DR PROSITE; PS50172; BRCT; 5.
SQ SEQUENCE 1056 AA; 119331 MW; 99CAF4EF4EE53ED6 CRC64;

alignment_scores:
Quality: 4059.50 Length: 1142
Ratio: 4.572 Gaps: 11
Percent Similarity: 77.758 Percent Identity: 70.928

alignment_block:
US-09-664-641-10 x Q9Z0W6 ..

Align seg 1/1 to: Q9Z0W6 from: 1 to: 1056

```
      33 CCGGAATTCCCGGGTCGACGATTTCGTCGAGTGCAAGACGGGCGGGCTGC 82
         ||||||   |||   :::   |||||||:::|||:::
       6 ProGluValProGluGluLeuPheArgGluValLys..............  17

      83 TGAGGGCGCCCCGCGGGTCTGGTCTGGGCCGGGGAACTCCGGGACGGCGC 132

      17 .................................................  17

     133 CGCAGCAGGGCTACGGATCTCTCGGCCCCTGCCCTTCAAATCACCTCTAG 182
                 :::   |||::::::   |||
      18 ..TyrTyrAlaValGlyAspIleAspPro.....................  26

     183 GCACAGGTTATTCAGCTTCTCAAGGCTGGAAAAGCGAAGGAAGTTTCCTA 232
             ||||||||||||||||||||||||||||||||||||||||||||||
      27 ...GlnValIleGlnLeuLeuLysAlaGlyLysAlaLysGluValSerTy  42

     233 CAATGCACTAGCCTCACACATAATCTCAGAGGATGGGGACAATCCAGAGG 282
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      42 rAsnAlaLeuAlaSerHisIleIleSerGluAspGlyAspAsnProGluV  59

     283 TGGGAGAAGCTCGGGAAGTCTTTGACTTACCTGTTGTAAAGCCTTCTTGG 332
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      59 alGlyGluAlaArgGluValPheAspLeuProValValLysProSerTrp  75

     333 GTGATTCTGTCCGTTCAGTGTGGAACTCTTCTGCCAGTAAATGGTTTTTC 382
         |||   ||||||||||||||||||:::|||||||||||||||||||||||
```


76 ValThrLeuSerValGlnCysGlyAlaLeuLeuProValAsnGlyPheSe 92
383 TCCAGAATCATGTCAGATTTTTTTTGGAATCACTGCCTGCCTTTCTCAGG 432
92 rProGluSerCysGlnIlePhePheGlyLeuThrAlaCysLeuSerGln. 108
433 GTGTTGATACAAGCTGGAGCTCTTTGTTGGAGTCTTCCAGAGCTCTCCCA 482
109 Val..... 109
483 AGGAGAGGTAGGGAAGGGAGCTTGTCCAGCAGAAGTTGGGAAGCACAGAG 532
109 109
533 ATCATCTGCCTTCTTCTGACCCGGTATTGATGCAGGCTGAGGCCTCTGTT 582
109 109
583 GTAATGTGCTGGGTGTCATCTGAAGACAGAAGTGCCCTGTGGGCTTTGGT 632
110 .....SerSerGluAspArgSerAlaLeuTrpAlaLeuVa 121
633 TACGTTCTATGGGGGAGATTGCCAGCTAACCCTCAATAAGAAATGCACGC 682
121 lThrPheHisGlyGlySerCysGlnLeuAsnLeuAsnLysLysCysThrH 138
683 ATTTGATTGTTCCAGAGCCAAAGGGGGGGAAATACGAATGTGCTTTAAAG 732
138 isLeuIleValProGluProLysGlyGluLysTyrGluArgAlaValLys 154
733 CGAGCAAGTATTAAAATTGTGACTCCTGACTGGGTTCTGGATTGCGTATC 782
155 ArgThrSerIleLysIleValThrProAspTrpValLeuAspCysValSe 171
783 AGAGAAAACCAAAAAGGACGAAGCATTTTATCATCCTCGTCTGATTATTT 832
171 rGluLysArgArgLysAspGluAlaPheTyrHisProArgLeuIleIleT 188
833 ATGAAGAGGAAGAAGAGGAAGAGGAAGAGGAGGAGGAAGTAGAAAATGAG 882
188 yrGluGluGluGluGluGluGluGluGly.....AspAsnGlu 201
883 GAACAAGATTCTCAGAATGAGGGTAGTACAGATGAGAAGTCAAGCCCTGC 932
202 GluGlnAspSerGlnAsnGluGlySerThr...GluLysSerServalAl 217
933 CAGCTCTCAAGAAGGGTCTCCTTCAGGTGACCAGCAGTTTTCACCTAAAT 982
217 aSerSerAlaValAlaSerProAla...GluGlnProCysSerProLysP 233
983 CCAACACTGAAAAATCTAAAGGGGAATTAATGTTTGATGATTCTTCAGAT 1032
233 roArgAlaGluValSerLysGlyGluLeuMetPheAspAspSerSerAsp 249
1033 TCATCACCGGAAAAACAGGAGAGAAATTTAAACTGGACCCCGGCCGAAGT 1082
250 SerSerProGluLysGlnGluArgSerLeuAsnTrpAlaProAlaGluAl 266
1083 CCCACAGTTAGCTGCAGCAAAACGCAGGCTGCCTCAGGGAAAGGAGCCTG 1132
266 aProProLeuAsnThrAlaGlnArgArgLeuProGlnGlyLysGlyProG 283
1133 GGTTGATTAACTTGTGTGCCAATGTCCCACCCGTCCCAGGTAACATTTTG 1182
283 lyLeuIleAsnLeuCysAlaAsnValProProValProGlyAspIleLeu 299
1183 CCCCCTGAGGTCCGGGGTAATTTAATGGCTGCTGGACAAAACCTCCAAAG 1232
300 ProProAspMetArgGlyAsnLeuMetAlaProGlyGlnAsnLeuGlnAs 316
1233 TTCTGAAAGATCAGAAATGATAGCTACCTGGAGTCCAGCTGTACGGACAC 1282
317 PheSerGluArgSerGluIleLeuGlyThrTrpSerProAlaValArgThrL 333
1283 TGAGGAATATTACTAATAATGCTGACATTCAGCAGATGAACCGGCCATCA 1332
333 euArgAsnIleThrAsnAsnAlaAspIleGlnGlnIleAsnArgProSer 349
1333 AATGTAGCACATATCTTACAGACTCTTTCAGCACCTACGAAAAATTTAGA 1382
350 AsnValAlaHisIleLeuGlnSerLeuSerAlaProThrLysSerLeuGl 366
1383 ACAGCAGGTGAATCACAGCCAGCAGGGACATACAAATGCCAATGCAGTGC 1432
366 uGlnGlnValAlaArgGlyGlnGlnGlyHisProAsnAlaSerAlaValL 383
1433 TGTTTAGCCAAGTGAAAGTGACTCCAGAGACACACATGCTACAGCAGCAG 1482
383 euPheGlyGlnAlaLysGlyAlaProGluThrHisValLeuGlnGlnHis 399
1483 CAGCAGGCCCAGCAGCAGCAGCAGCAGCACCCGGTTTTACACCTTCAGCC 1532
400 HisProProGlnGlnProGlnGlnGlnHisProAlaLeuHisLeuGlnPr 416
1533 CCAGCAGATAATGCAGCTCCAGCAGCAGCAGCAGCAGATCTCTCAG. 1581
416 o...GlnIleMetGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnG 432
1582 .....CAACCTTACCCCCAGCAGCCGCCGCATCCATTTTCA... 1617
432 lnGlnGlnProGlnProTyrProGlnProProSerHisGlnPheProGln 448
1617 1617
449 GlnValHisGlnHisGlnPheSerGlnGlnGlnLeuGlnPheProGlnGl 465
1617 1617
465 nProLeuHisProGlnGlnGlnLeuHisArgProGlnGlnGlnLeuGlnP 482
1617 1617
482 roPheGlnGlnGlnHisAlaLeuGlnGlnGlnLeuHisGlnLeuGlnGln 498
1617 1617
499 GlnGlnLeuGlnHisHisGlnLeuAlaGlnLeuGlnGlnGlnGlnGlnGl 515
1618 .....CAGCAGCAGCAGCAGCAACAGCAG..... 1641
515 nGlnHisAsnLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnLeuGlnArgL 532
1641 1641
532 euGlnGlnGlnGlnGlnMetGlnAsnGlnAlaAlaHisLeuSerGlnAla 548
1641 1641
549 SerGlnAlaLeuGlnHisGlnValLeuProGlnGlnProLeuGlnLeuSe 565
1642 .....CCACCACCATCGCCTCAGCAGCATCAGCTTTTTGGACATG 1681
565 rLeuGlnProProProGlnGlnGlnGlnGlnGlnGlnLeuPheGlyHisA 582
1682 ATCCAGCAGTGGAGATTCCAGAAGAAGGCTTCTTATTGGGATGTGTGTTT 1731
582 spProAlaValGluIleProGluGluSerPheLeuLeuGlyCysValPhe 598
1732 GCAATTGCGGATTATCCAGAGCAGATGTCTGATAAGCAACTGCTGGCCAC 1781
599 AlaIleAlaAspTyrProGluGlnMetSerAspLysGlnLeuLeuAlaTh 615
1782 CTGGAAAAGGATAATCCAGGCACATGGCGGCACTGTTGACCCCACCTTCA 1831
615 rTrpLysArgIleIleGlnAlaHisGlyGlyThrValAspProThrPheT 632

```
1832 CGAGTCGATGCACGCACCTTCTCTGTGAGAGTCAAGTCAGCAGCGCGTAT 1881
      |||||||||||||||||||||||||   ||||||||||||||||   |||
 632 hrSerArgCysThrHisLeuLeuCysAlaSerGlnValSerSerMetTyr 648

1882 GCACAGGCAATAAGAGAAAGAAAGAGATGTGTTACTGCACACTGGTTAAA 1931
      :::|||||:::|||||||||||||||||||||||||||||||||||||||
 649 ThrGlnAlaLeuArgGluArgLysArgCysValThrAlaHisTrpLeuAs 665

1932 CACAGTCTTAAAAAAAAAAAAAAATGGTACCGCCGCACCGAGCCCTTCACT 1981
      |||||||||||||||||||||:::::|||||||||||||||||||||||||
 665 nThrValLeuLysLysLysLysLeuMetProProHisArgAlaLeuHisP 682

1982 TCCCAGTGGCCTTCCCACCAGGAGGAAAGCCATGTTCACAGCATATTATT 2031
      |||||||||||||||||||||||||||||||||||||||||||||||||
 682 heProValAlaPheProProGlyGlyLysProCysSerGlnHisIleIle 698

2032 TCTGTGACTGGATTTGTTGATAGTGACAGAGATGACCTAAAATTAATGGC 2081
      |||||||||||||||||||||:::|||||||||||||||||||||||||
 699 SerValThrGlyPheValAspAsnAspArgAspAspLeuLysLeuMetAl 715

2082 TTATTTGGCAGGTGCCAAATATACGGGTTATCTATGCCGCAGCAACACAG 2131
      |||||||||||||||||||||||||||||||||||||||||||||||||
 715 aTyrLeuAlaGlyAlaLysTyrThrGlyTyrLeuCysArgSerAsnThrV 732

2132 TCCTCATCTGTAAAGAACCAACTGGTTTAAAGTATGAAAAAGCCAAAGAG 2181
      |||||||||||||||||||:::|||||||||||||||||||||||||||
 732 alLeuIleCysLysGluProSerGlyLeuLysTyrGluLysAlaLysGlu 748

2182 TGGAGGATACCCTGTGTCAACGCCCAGTGGCTTGGCGACATTCTTCTGGG 2231
      |||||||||||||||||||||||||||||||||||||||||||||||||
 749 TrpArgIleProCysValAsnAlaGlnTrpLeuGlyAspIleLeuLeuGl 765

2232 AAACTTTGAGGCACTGAGGCAGATTCAGTATAGTCGCTACACGGCATTCA 2281
      |||||||||||||||||||||:::|||||||||||||||||||||||||:
 765 yAsnPheGluAlaLeuArgGlnValGlnTyrSerArgTyrThrAlaPheA 782

2282 GTCTGCAGGATCCATTTGCCCCTACCCAGCATTTAGTTTTAAATCTTTTA 2331
      :::::   ||||||||||:::|||||   |||||||||||||:::|||||
 782 snMetProAspProPheValProThrProHisLeuValLeuGlyLeuLeu 798

2332 GATGCTTGGAGAGTTCCCTTAAAAGTGTCTGCAGAGTTGTTGATGAGTAT 2381
      ||||||||||:::|||:::|||||||:::|||||||||||||||:::::
 799 AspAlaTrpArgThrProValLysValThrAlaGluLeuLeuMetGlyVa 815

2382 AAGACTACCTCCCAAACTGAAACAGAATGAAGTAGCTAATGTCCAGCCTT 2431
      :||||   |||||||||||   ||||||||||||||:::|||   |
 815 lArgLeu***ProLysLeuLysProAsnGluValAlaAsnIleGlnLeuS 832

2432 CTTCCCAAAGAGCCAGAATTGAAGACGTACCACCTCCCACTAAAAAGCTA 2481
      |||||:::|||||||||||||||||||:::|||||||||||||||||||
 832 erSerLysArgAlaArgIleGluAspLeuProProProThrLysLysLeu 848

2482 ACTCCAGAATTGACCCCTTTTGTGCTTTTCACTGGATTCGAGCCTGTCCA 2531
      ||||||||||||||||||:::|||||||||||||||||||||||||||||
 849 ThrProGluLeuThrProLeuValLeuPheThrGlyPheGluProValGl 865

2532 GGTTCAACAGTATATTAAGAAGCTCTACATTCTTGGTGGAGAGGTTGCGG 2581
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 865 nValGlnGlnTyrIleLysLysLeuTyrIleLeuGlyGlyGluValAlaG 882

2582 AGTCTGCACAGAAGTGCACACACCTCATTGCCAGCAAAGTGACTCGCACC 2631
      ||   ::::::|||||||||||||||||||||||||||||||||||||||
 882 luCysThrLysLysCysThrHisLeuIleAlaSerLysValThrArgThr 898

2632 GTGAAGTTCCTGACGGCGATTTCTGTCGTGAAGCACATAGTGACGCCAGA 2681
      ||||||||||||||||||||||||||||||||||||||||||||||||::
 899 ValLysPheLeuThrAlaIleSerValValLysHisIleValThrProAs 915

2682 GTGGCTGGAAGAATGCTTCAGGTGTCAGAAGTTCATTGATGAGCAGAACT 2731
      :|||||||||||||||||:::   |||   ||||||||||||||||||||
 915 pTrpLeuGluGluCysPheLysArgGlnThrPheIleAspGluGlnAsnT 932

2732 ACATTCTCCGAGATGCTGAGGCAGAAGTACTTTTCTCTTTCAGCTTGGAA 2781
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 932 yrIleLeuArgAspAlaGluAlaGluValLeuPheSerPheSerLeuGlu 948

2782 GAATCCTTAAAACGGGCACACGTTTCTCCACTCTTTAAGGCAAAATATTT 2831
      |||||||||||||||||||||||||||||:::|||||:::||||||||
 949 GluSerLeuLysArgAlaHisValSerProPhePheLysThrLysTyrPh 965

2832 TTACATCACACCTGGAATCTGCCCAAGTCTTTCCACTATGAAGGCAATCG 2881
      |||||||||||||||||||||||||||||||:::|||||||||||||||
 965 eTyrIleThrProGlyIleCysProSerLeuAlaThrMetLysAlaIleV 982

2882 TAGAGTGTGCAGGAGGAAAGGTGTTATCCAAGCAGCCATCTTTCCGGAAG 2931
      |||||||||||||||||||||||||||:::|||||||||||||||||||
 982 alGluCysAlaGlyGlyLysValLeuAlaLysGlnProSerPheArgLys 998

2932 CTCATGGAGCACAAGCAGAACTCGAGTTTGTCGGAAATAATTTTAATATC 2981
      |||||||||||||||||||:::|||||||||||||||||||||||||||
 999 LeuMetGluHisLysGlnAsnLysSerLeuSerGluIleIleLeuIleSe 1015

2982 CTGTGAAAATGACCTTCATTTATGCCGAGAATATTTTGCCAGAGGCATAG 3031
      ||||||||||||||||||||||||||||||||||||||||||||||||||
1015 rCysGluAsnAspLeuHisLeuCysArgGluTyrPheAlaArgGlyIleA 1032

3032 ATGTTCACAATGCAGAGTTCGTTCTGACTGGAGTGCTCACTCAAACGCTG 3081
      ||||||||||||||||||||||||||||||||||||||||||||||||||
1032 spValHisAsnAlaGluPheValLeuThrGlyValLeuThrGlnThrLeu 1048

3082 GACTATGAATCATATAAGTTTAAC 3105
      |||||||||||||||||||||||
1049 AspTyrGluSerTyrLysPheAsn 1056
```

seq_name: sp_human:O15404

```
seq_documentation_block:
ID   O15404      PRELIMINARY;      PRT;   744 AA.
AC   O15404;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CAGF28 (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=97369492; PubMed=9225980;
RA   Margolis R.L., Abraham M.R., Gatchell S.B., Li S.H., Kidwai A.S.,
RA   Breschel T.S., Stine O.C., Callahan C., McInnis M.G., Ross C.A.;
RT   "cDNAs with long CAG trinucleotide repeats from human brain.";
RL   Hum. Genet. 100:114-122(1997).
DR   EMBL; U80735; AAB91434.1; -.
DR   InterPro; IPR001357; BRCT.
DR   Pfam; PF00533; BRCT; 4.
DR   SMART; SM00292; BRCT; 4.
DR   PROSITE; PS50172; BRCT; 2.
FT   NON_TER       1      1
SQ   SEQUENCE   744 AA;  83747 MW;  8FB1D5EC267A70E0 CRC64;


alignment_scores:
           Quality: 3841.00              Length:    745
             Ratio:  5.163                 Gaps:      1
 Percent Similarity:  99.866   Percent Identity: 99.463

alignment_block:
 US-09-664-641-10 x O15404    ..

 Align seg 1/1  to: O15404  from: 1  to: 744

 874 GAAAATGAGGAACAAGATTCTCAGAATGAGGGTAGTACAGATGAGAAGTC 923
      ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
   1 GluAsnGluGluGlnAspSerGlnAsnGluGlySerThrAspGluLysSe 17

 924 AAGCCCTGCCAGCTCTCAAGAAGGGTCTCCTTCAGGTGACCAGCAGTTTT 973
  17 rSerProAlaSerSerGlnGluGlySerProSerGlyAspGlnGlnPheS 34

 974 CACCTAAATCCAACACTGAAAAATCTAAAGGGGAATTAATGTTTGATGAT 1023
  34 erProLysSerAsnThrGluLysSerLysGlyGluLeuMetPheAspAsp 50

1024 TCTTCAGATTCATCACCGGAAAAACAGGAGAGAAATTTAAACTGGACCCC 1073
  51 SerSerAspSerSerProGluLysGlnGluArgAsnLeuAsnTrpThrPr 67

1074 GGCCGAAGTCCCACAGTTAGCTGCAGCAAAACGCAGGCTGCCTCAGGGAA 1123
  67 oAlaGluValProGlnLeuAlaAlaAlaLysArgArgLeuProGlnGlyL 84

1124 AGGAGCCTGGGTTGATTAACTTGTGTGCCAATGTCCCACCCGTCCCAGGT 1173
  84 ysGluProGlyLeuIleAsnLeuCysAlaAsnValProProValProGly 100

1174 AACATTTTGCCCCCTGAGGTCCGGGGTAATTTAATGGCTGCTGGACAAAA 1223
 101 AsnIleLeuProProGluValArgGlyAsnLeuMetAlaAlaGlyGlnAs 117

1224 CCTCCAAAGTTCTGAAAGATCAGAAATGATAGCTACCTGGAGTCCAGCTG 1273
 117 nLeuGlnSerSerGluArgSerGluMetIleAlaThrTrpSerProAlaV 134

1274 TACGGACACTGAGGAATATTACTAATAATGCTGACATTCAGCAGATGAAC 1323
 134 alArgThrLeuArgAsnIleThrAsnAsnAlaAspIleGlnGlnMetAsn 150

1324 CGGCCATCAAATGTAGCACATATCTTACAGACTCTTTCAGCACCTACGAA 1373
 151 ArgProSerAsnValAlaHisIleLeuGlnThrLeuSerAlaProThrLy 167

1374 AAATTTAGAACAGCAGGTGAATCACAGCCAGCAGGGACATACAAATGCCA 1423
 167 sAsnLeuGluGlnGlnValAsnHisSerGlnGlnGlyHisThrAsnAlaA 184

1424 ATGCAGTGCTGTTTAGCCAAGTGAAAGTGACTCCAGAGACACACATGCTA 1473
 184 snAlaValLeuPheSerGlnValLysValThrProGluThrHisMetLeu 200

1474 CAGCAGCAGCAGCAGGCCCAGCAGCAGCAGCAGCACCCGGTTTTACA 1523
 201 GlnGlnGlnGlnGlnAlaGlnGlnGlnGlnGlnHisProValLeuHi 217

1524 CCTTCAGCCCCAGCAGATAATGCAGCTCCAGCAGCAGCAGCAGCAGA 1573
 217 sLeuGlnProGlnGlnIleMetGlnLeuGlnGlnGlnGlnGlnGlnI 234

1574 TCTCTCAGCAACCTTACCCCCAGCAGCCGCCGCATCCATTTTCACAGCAG 1623
 234 leSerGlnGlnProTyrProGlnGlnProProHisProPheSerGlnGln 250

1624 CAGCAGCAGCAACAGCAGCCACCACCATCGCCTCAGCAGCATCAGCTTTT 1673
 251 GlnGlnGlnGlnGlnProProProSerProGlnGlnHisGlnLeuPh 267

1674 TGGACATGATCCAGCAGTGGAGATTCCAGAAGAAGGCTTCTTATTGGGAT 1723
 267 eGlyHisAspProAlaValGluIleProGluGluGlyPheLeuLeuGlyC 284

1724 GTGTGTTTGCAATTGCGGATTATCCAGAGCAGATGTCTGATAAGCAACTG 1773
 284 ysValPheAlaIleAlaAspTyrProGluGlnMetSerAspLysGlnLeu 300

1774 CTGGCCACCTGGAAAAGGATAATCCAGGCACATGGCGGCACTGTTGACCC 1823
      LeuAlaThrTrpLysArgIleIleGlnAlaHisGlyGlyThrValAsp.P 317

1824 CACCTTCACGAGTCGATGCACGCACCTTCTCTGTGAGAGTCAAGTCAGCA 1873
 317 roProSerArgValAspAlaArgThrPheSerValArgValLysSerAla 333

1874 GCGCGTAT.GCACAGGCAATAAGAGAAAGAAAGAGATGTGTTACTGCACA 1922
 334 AlaArgIleAlaGlnAlaIleArgGluArgLysArgCysValThrAlaHi 350

1923 CTGGTTAAACACAGTCTTAAAAAAAAAAAAAAATGGTACCGCCGCACCGAG 1972
 350 sTrpLeuAsnThrValLeuLysLysLysLysMetValProProHisArgA 367

1973 CCCTTCACTTCCCAGTGGCCTTCCCACCAGGAGGAAAGCCATGTTCACAG 2022
 367 laLeuHisPheProValAlaPheProProGlyGlyLysProCysSerGln 383

2023 CATATTATTTCTGTGACTGGATTTGTTGATAGTGACAGAGATGACCTAAA 2072
 384 HisIleIleSerValThrGlyPheValAspSerAspArgAspAspLeuLy 400

2073 ATTAATGGCTTATTTGGCAGGTGCCAAATATACGGGTTATCTATGCCGCA 2122
 400 sLeuMetAlaTyrLeuAlaGlyAlaLysTyrThrGlyTyrLeuCysArgS 417

2123 GCAACACAGTCCTCATCTGTAAAGAACCAACTGGTTTAAAGTATGAAAAA 2172
 417 erAsnThrValLeuIleCysLysGluProThrGlyLeuLysTyrGluLys 433

2173 GCCAAAGAGTGGAGGATACCCTGTGTCAACGCCCAGTGGCTTGGCGACAT 2222
 434 AlaLysGluTrpArgIleProCysValAsnAlaGlnTrpLeuGlyAspIl 450

2223 TCTTCTGGGAAACTTTGAGGCACTGAGGCAGATTCAGTATAGTCGCTACA 2272
 450 eLeuLeuGlyAsnPheGluAlaLeuArgGlnIleGlnTyrSerArgTyrT 467

2273 CGGCATTCAGTCTGCAGGATCCATTTGCCCCTACCCAGCATTTAGTTTTA 2322
 467 hrAlaPheSerLeuGlnAspProPheAlaProThrGlnHisLeuValLeu 483

2323 AATCTTTTAGATGCTTGGAGAGTTCCCTTAAAAGTGTCTGCAGAGTTGTT 2372
 484 AsnLeuLeuAspAlaTrpArgValProLeuLysValSerAlaGluLeuLe 500

2373 GATGAGTATAAGACTACCTCCCAAACTGAAACAGAATGAAGTAGCTAATG 2422
 500 uMetSerIleArgLeuProProLysLeuLysGlnAsnGluValAlaAsnV 517

2423 TCCAGCCTTCTTCCCAAAGAGCCAGAATTGAAGACGTACCACCTCCCACT 2472
 517 alGlnProSerSerGlnArgAlaArgIleGluAspValProProProThr 533

2473 AAAAAGCTAACTCCAGAATTGACCCCTTTTGTGCTTTTCACTGGATTCGA 2522
 534 LysLysLeuThrProGluLeuThrProPheValLeuPheThrGlyPheGl 550

2523 GCCTGTCCAGGTTCAACAGTATATTAAGAAGCTCTACATTCTTGGTGGAG 2572
 550 uProValGlnValGlnGlnTyrIleLysLysLeuTyrIleLeuGlyGlyG 567

2573 AGGTTGCGGAGTCTGCACAGAAGTGCACACACCTCATTGCCAGCAAAGTG 2622
 567 luValAlaGluSerAlaGlnLysCysThrHisLeuIleAlaSerLysVal 583

2623 ACTCGCACCGTGAAGTTCCTGACGGCGATTTCTGTCGTGAAGCACATAGT 2672
 584 ThrArgThrLeuLysPheLeuAlaAlaIleSerValValLysHisIleVa 600

2673 GACGCCAGAGTGGCTGGAAGAATGCTTCAGGTGTCAGAAGTTCATTGATG 2722
 600 lThrProGluTrpLeuGluGluCysPheArgCysGlnLysPheIleAspG 617
```

```
2723 AGCAGAACTACATTCTCCGAGATGCTGAGGCAGAAGTACTTTTCTCTTTC 2772
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 617 luGlnAsnTyrIleLeuArgAspAlaGluAlaGluValLeuPheSerPhe 633

2773 AGCTTGGAAGAATCCTTAAAACGGGCACACGTTTCTCCACTCTTTAAGGC 2822
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 634 SerLeuGluGluSerLeuLysArgAlaHisValSerProLeuPheLysAl 650

2823 AAAATATTTTTACATCACACCTGGAATCTGCCCAAGTCTTTCCACTATGA 2872
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 650 aLysTyrPheTyrIleThrProGlyIleCysProSerLeuSerThrMetL 667

2873 AGGCAATCGTAGAGTGTGCAGGAGGAAAGGTGTTATCCAAGCAGCCATCT 2922
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 667 ysAlaIleValGluCysAlaGlyGlyLysValLeuSerLysGlnProSer 683

2923 TTCCGGAAGCTCATGGAGCACAAGCAGAACTCGAGTTTGTCGGAAATAAT 2972
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 684 PheArgLysLeuMetGluHisLysGlnAsnSerSerLeuSerGluIleIl 700

2973 TTTAATATCCTGTGAAAATGACCTTCATTTATGCCGAGAATATTTTGCCA 3022
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 700 eLeuIleSerCysGluAsnAspLeuHisLeuCysArgGluTyrPheAlaA 717

3023 GAGGCATAGATGTTCACAATGCAGAGTTCGTTCTGACTGGAGTGCTCACT 3072
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 717 rgGlyIleAspValHisAsnAlaGluPheValLeuThrGlyValLeuThr 733

3073 CAAACGCTGGACTATGAATCATATAAGTTTAAC 3105
     |||||||||||||||||||||||||||||||||
 734 GlnThrLeuAspTyrGluSerTyrLysPheAsn 744

seq_name: sp_vertebrate:Q90WJ3

seq_documentation_block:
ID   Q90WJ3      PRELIMINARY;      PRT;  1256 AA.
AC   Q90WJ3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SWIFT.
GN   K14.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE=21260044; PubMed=11359898;
RA   Shimizu K., Bourillot P.Y., Nielsen S.J., Zorn A.M., Gurdon J.B.;
RT   "Swift Is a Novel BRCT Domain Coactivator of Smad2 in Transforming
RT   Growth Factor beta Signaling.";
RL   Mol. Cell. Biol. 21:3901-3912(2001).
DR   EMBL; AF172855; AAK55123.1; -.
SQ   SEQUENCE   1256 AA;  144187 MW;  E7795B12C1A42DE1 CRC64;


alignment_scores:
            Quality: 3425.50              Length:   1332
              Ratio:  4.112                 Gaps:     20
 Percent Similarity:  62.538   Percent Identity: 53.829

alignment_block:
 US-09-664-641-10 x Q90WJ3   ..

 Align seg 1/1  to: Q90WJ3  from: 1  to: 1256

    165 CCTTCAAATCACCTCTAGGCACAGGTTATTCAGCTTCTCAAGGCTGGAAA 214
        |||   :::   |||       :::|||:::|||:::|||||||||||||||
     20 ProValGlyAspLeuAspProLysValValGlnMetLeuLysAlaGlyLy 36

    215 AGCGAAGGAAGTTTCCTACAATGCACTAGCCTCACACATAATCTCAGAGG 264
        ||||||||||||||||||||||||||||:::|||||||||:::||||
     36 sAlaLysGluValSerTyrAsnAlaLeuAlaThrHisIleIleAlaGluA 53

    265 ATGGGGACAATCCAGAGGTGGGAGAAGCTCGGGAAGTCTTTGACTTACCT 314
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
     53 spGlyAspAsnProGluValGlyGluAlaArgGluValPheAspLeuPro 69

    315 GTTGTAAAGCCTTCTTGGGTGATTCTGTCCGTTCAGTGTGGAACTCTTCT 364
        ||||||||||||||||||||||||||||||:::|||||:::   ||
     70 ValValLysProSerTrpValIleLeuSerValArgCysGlyValArgLe 86

    365 GCCAGTAAATGGTTTTTCTCCAGAATCATGTCAGATTTTTTTTGGAATCA 414
        ||||   ||||||||||||||||||||   ||||||||||||||:::|
     86 uProGluAsnGlyPheSerProGluSerGlyGlnIlePhePheGlyValT 103

    415 CTGCCTGCCTTTCTCAGGGTGTTGATACAAGCTGGAGCTCTTTGTTGGAG 464
        ||||||||||||||||| |||
    103 hrAlaCysLeuSerGln.Val..................................... 109

    465 TCTTCCAGAGCTCTCCCAAGGAGAGGTAGGGAAGGGAGCTTGTCCAGCAG 514

    109 ..... 109

    515 AAGTTGGGAAGCACAGAGATCATCTGCCTTCTTCTGACCCGGTATTGATG 564

    109 ..... 109

    565 CAGGCTGAGGCCTCTGTTGTAATGTGCTGGGTGTCATCTGAAGACAGAAG 614
                                    |||   :::||||||::
    110 ..............................SerProAspAspArgAs 115

    615 TGCCCTGTGGGCTTTGGTTACGTTCTATGGGGGAGATTGCCAGCTAACCC 664
        ::::|||||||||||||:::|||||||||||||||||||||||||||:::|
    115 nSerLeuTrpAlaLeuThrThrPheTyrGlyGlyAspCysGlnLeuSerL 132

    665 TCAATAAGAAATGCACGCATTTGATTGTTCCAGAGCCAAAGGGGGGGAAA 714
        ||||||||||||||||||||||||||||||||||||||||||||:::|||
    132 euAsnLysLysCysThrHisLeuIleValProGluProLysGlyAsnLys 148

    715 TACGAATGTGCTTTAAAGCGAGCAAGTATTAAAATTGTGACTCCTGACTG 764
        ||||||   |||::::::|||:::|||||||||||||||||||||||||
    149 TyrGluTyrAlaPheGlnArgGlySerIleLysIleValThrProAspTr 165

    765 GGTTCTGGATTGCGTATCAGAGAAAACCAAAAAGGACGAAGCATTTTATC 814
        |||||||||   ||||||||||||||||||||||||||||||:::||||
    165 pValLeuAspSerValSerGluLysThrLysLysAspGluAlaLeuTyrH 182

    815 ATCCTCGTCTGATTATTTATGAAGAGGAAGAAGAGGAAGAGGAAGAGGAG 864
        ||||||||||||:::|||      ||||||||||||||||||||||||
    182 isProArgLeuIleValTyr......GluGluGluGluGluGluGluGlu 196

    865 GAGGAAGTAGAAAATGAGGAACAAGATTCTCAGAATGAGGGTAGTACAGA 914
        ||||||:::|||||||||||||||::|||||||||   ::::::   ||
    197 GluGluAlaGluAsnGluGluGlnGluSerGlnAsnGlySerAspGluAs 213

    915 TGAG.........AAGTCAAGCCCTGCCAGCTCTCAAGAAGGGTCTCCTT 955
        |:::         ||||||||||||:::::|||::::::|||||||||
    213 pLysTyrSerSerLysSerSerProValThrSerArgAspGlySerProV 230

    956 CAGGTGACCAGCAGTTTTCACCTAAA...TCCAACACTGAAAAATCTAAA 1002
            :::::::::   |||||||||   |||:::::::::|||   |||
    230 alHisAsnArgGluSerSerProLysArgSerSerSerAspLysIleLys 246

   1003 GGGGAATTAATGTTTGATGATTCTTCAGATTCATCACCGGAAAAACAGGA 1052
        :::|||||||||||||||||||||||||||||||||||||||||||||||
    247 SerGluLeuMetPheAspAspSerSerAspSerSerProGluLysGlnGl 263

   1053 GAGAAATTTAAACTGGACCCCGGCCGAAGTCCCACAGTTAGCTGCAGCAA 1102
        ||||||||||||||||||||||||||||||||||||:::::::::||||
    263 uArgAsnLeuAsnTrpThrProAlaGluValProGlnMetSerThrAlaL 280

   1103 AACGCAGGCTGCCTCAGGGAAAGGAGCCTGGGTTGATTAACTTGTGTGCC 1152
```

OM of: US-09-664-641-10 to: SPTREMBL_19:* out_format : pfs

Date: Jun 4, 2002 5:49 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=x1p
-Q=/cgn2_1/USPTO_spool/US09664641/runat_04062002_110933_9250/app_query.fasta_1.3697
-DB=SPTREMBL_19 -QFMT=fastan -SUFFIX=rspt -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09664641_@CGN1_1_312 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-664-641-10
Query length: 3580
Database: SPTREMBL_19:*
Database sequences: 562222
Database length: 172994929
Search time (sec): 230.300000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
sp_rodent:Q9Z0W6	+	4059.50	5875.12	0.0	1056	! Q9z0w6 mus musculus (mouse). p
sp_human:O15404	+	3841.00	5561.10	0.0	744	! O15404 homo sapiens (human). ca
sp_vertebrate:Q90WJ3	+	3425.50	4951.40	1.5e-268	1256	! Q90wj3 xenopus laevis (africar
sp_human:Q96HP2	+	2039.00	2947.67	1.9e-156	391	! Q96hp2 homo sapiens (human). un
sp_invertebrate:Q9VUB6	+	852.00	1205.56	4.6e-60	1798	! Q9vub6 drosophila melanogast
sp_invertebrate:Q94046	+	524.00	734.07	1.4e-33	1074	! Q94046 caenorhabditis elegar
sp_invertebrate:Q9U370	+	524.00	734.05	1.4e-33	1076	! Q9u370 caenorhabditis elegar
sp_human:Q14676	+	342.50	463.12	8.9e-19	2089	! Q14676 homo sapiens (human). k
sp_human:Q96QC2	+	340.50	460.21	1.3e-18	2090	! Q96qc2 homo sapiens (human). k
sp_plant:Q9LVF7	+	316.50	432.68	8.9e-17	1041	! Q9lvf7 arabidopsis thaliana (m
sp_rodent:Q91VX3	+	283.50	382.38	4.5e-14	1296	! Q9lvx3 mus musculus (mouse). s
sp_plant:Q9ZR18	+	265.50	361.78	1.1e-12	765	! Q9zr18 arabidopsis thaliana (mc
sp_human:Q9UEB9	+	263.50	352.22	2.0e-12	1435	! Q9ueb9 homo sapiens (human). d
sp_human:Q92547	+	263.50	351.40	2.0e-12	1550	! Q92547 homo sapiens (human). k
sp_fungi:Q9P5J5	+	229.00	307.74	1.0e-09	838	! Q9p5j5 neurospora crassa. conse
sp_invertebrate:Q9VFR7	+	225.50	303.46	1.9e-09	776	! Q9vfr7 drosophila melanogaste
sp_invertebrate:Q24163	+	220.50	291.76	5.5e-09	1180	! Q24163 drosophila melanogast
sp_invertebrate:Q9VED3	+	220.50	291.71	5.6e-09	1186	! Q9ved3 drosophila melanogast
sp_invertebrate:O44011	+	219.50	288.08	7.2e-09	1457	! O44011 dictyostelium discoid
sp_invertebrate:Q9W4E8	+	216.00	293.94	9.6e-09	517	! Q9w4e8 drosophila melanogaste
sp_plant:Q93VQ3	+	215.50	283.70	1.4e-08	1272	! Q93vq3 oryza sativa (rice). hy
sp_invertebrate:O76853	+	214.00	293.23	1.3e-08	420	! O76853 dictyostelium discoide
sp_fungi:P79025	+	214.00	291.75	1.4e-08	483	! P79025 emericella nidulans (asp
sp_virus:Q98148	+	214.00	282.47	1.9e-08	1162	! Q98148 kaposi's sarcoma-associ
sp_invertebrate:Q9VP17	+	213.50	274.60	2.6e-08	2285	! Q9vp17 drosophila melanogast
sp_vertebrate:Q90YI2	+	212.50	275.54	2.9e-08	1822	! Q90yi2 fugu rubripes (japanese
sp_invertebrate:Q9Y0C9	+	212.00	283.02	2.4e-08	838	! Q9y0c9 dictyostelium discoide
sp_invertebrate:O18346	+	210.50	278.87	3.4e-08	1010	! O18346 drosophila hydei (fru
sp_virus:Q9QR71	+	210.00	276.96	3.9e-08	1129	! Q9qr71 kaposi's sarcoma-associ
sp_invertebrate:Q9VMS2	+	208.50	266.64	6.7e-08	2439	! Q9vms2 drosophila melanogast
sp_invertebrate:Q95ZG5	+	207.00	262.29	9.6e-08	2994	! Q95zg5 dictyostelium discoid
sp_invertebrate:Q9V7C5	+	206.50	273.98	6.9e-08	925	! Q9v7c5 drosophila melanogaste
sp_invertebrate:Q95YM8	+	206.50	268.20	8.4e-08	1598	! Q95ym8 apis mellifera (honey
sp_invertebrate:Q17641	-	205.50	288.88	4.8e-08	197	! Q17641 caenorhabditis elegans
sp_invertebrate:Q9VN59	+	205.00	271.70	9.2e-08	934	! Q9vn59 drosophila melanogaste
sp_invertebrate:Q94485	+	204.50	274.64	8.9e-08	660	! Q94485 dictyostelium discoide
sp_invertebrate:Q9N8T2	-	203.00	255.01	2.1e-07	3439	! Q9n8t2 trypanosoma brucei. p
sp_human:O15410	+	202.50	271.86	1.3e-07	652	! O15410 homo sapiens (human). ca
sp_human:O75557	+	202.50	259.89	1.9e-07	2023	! O75557 homo sapiens (human). c
sp_human:Q9UND7	+	202.50	259.89	1.9e-07	2023	! Q9und7 homo sapiens (human). d
sp_virus:Q9DUN0	+	201.00	265.41	2.0e-07	976	! Q9dun0 kaposi's sarcoma-asso
sp_rodent:Q64507	-	200.00	279.24	1.4e-07	230	! Q64507 mus musculus (mouse).
sp_human:Q9BT44	+	200.00	268.27	2.0e-07	649	! Q9bt44 homo sapiens (human).
sp_virus:O40947	+	200.00	262.80	2.5e-07	1089	! O40947 kaposi's sarcoma-ass
sp_invertebrate:Q9U9S7	+	200.00	255.74	3.1e-07	2123	! Q9u9s7 dictyostelium disc

seq_name: sp_rodent:Q9Z0W6

seq_documentation_block:
ID Q9Z0W6 PRELIMINARY; PRT; 1056 AA.
AC Q9Z0W6;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PAX TRANSCRIPTION ACTIVATION DOMAIN INTERACTING PROTEIN PTIP.
GN PAXIP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB;
RX MEDLINE=20368635; PubMed=10908331;
RA Lechner M.S., Levitan I., Dressler G.R.;
RT "PTIP, a novel BRCT domain-containing protein interacts with Pax2 and
RT is associated with active chromatin.";
RL Nucleic Acids Res. 28:2741-2751(2000).
DR EMBL; AF104261; AAD17923.1; -.
DR MGD; MGI:1890430; Paxip1.
DR InterPro; IPR001357; BRCT.
DR Pfam; PF00533; BRCT; 6.
DR SMART; SM00292; BRCT; 5.
DR PROSITE; PS50172; BRCT; 5.
SQ SEQUENCE 1056 AA; 119331 MW; 99CAF4EF4EE53ED6 CRC64;

alignment_scores:
Quality: 4059.50 Length: 1142
Ratio: 4.572 Gaps: 11
Percent Similarity: 77.758 Percent Identity: 70.928

alignment_block:
US-09-664-641-10 x Q9Z0W6 ..

Align seg 1/1 to: Q9Z0W6 from: 1 to: 1056

```
     33 CCGGAATTCCCGGGTCGACGATTTCGTCGAGTGCAAGACGGGCGGGCTGC 82
        ||||||   |||   :::   ||||||:::|||:::
      6 ProGluValProGluGluLeuPheArgGluValLys..... 17

     83 TGAGGGCGCCCCGCGGGTCTGGTCTGGGCCGGGGAACTCCGGGACGGCGC 132

     17 ..................................................... 17

    133 CGCAGCAGGGCTACGGATCTCTCGGCCCCTGCCCTTCAAATCACCTCTAG 182
              :::   |||::::::   |||
     18 ..TyrTyrAlaValGlyAspIleAspPro..... 26

    183 GCACAGGTTATTCAGCTTCTCAAGGCTGGAAAAGCGAAGGAAGTTTCCTA 232
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     27 ...GlnValIleGlnLeuLeuLysAlaGlyLysAlaLysGluValSerTy 42

    233 CAATGCACTAGCCTCACACATAATCTCAGAGGATGGGGACAATCCAGAGG 282
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     42 rAsnAlaLeuAlaSerHisIleIleSerGluAspGlyAspAsnProGluV 59

    283 TGGGAGAAGCTCGGGAAGTCTTTGACTTACCTGTTGTAAAGCCTTCTTGG 332
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     59 alGlyGluAlaArgGluValPheAspLeuProValValLysProSerTrp 75

    333 GTGATTCTGTCCGTTCAGTGTGGAACTCTTCTGCCAGTAAATGGTTTTTC 382
        |||   ||||||||||||||||||:::|||||||||||||||||||||||
```

```
 76 ValThrLeuSerValGlnCysGlyAlaLeuLeuProValAsnGlyPheSe 92

383 TCCAGAATCATGTCAGATTTTTTTTGGAATCACTGCCTGCCTTTCTCAGG 432
    ||||||||||||||||||||||||||||:::||||||||||||||||||||
 92 rProGluSerCysGlnIlePhePheGlyLeuThrAlaCysLeuSerGln. 108

433 GTGTTGATACAAGCTGGAGCTCTTTGTTGGAGTCTTCCAGAGCTCTCCCA 482
    |||
109 Val..................................................  109

483 AGGAGAGGTAGGGAAGGGAGCTTGTCCAGCAGAAGTTGGGAAGCACAGAG 532

109 ..................................................... 109

533 ATCATCTGCCTTCTTCTGACCCGGTATTGATGCAGGCTGAGGCCTCTGTT 582

109 ..................................................... 109

583 GTAATGTGCTGGGTGTCATCTGAAGACAGAAGTGCCCTGTGGGCTTTGGT 632
     }            ||||||||||||||||||||||||||||||||||||||||||
110 ..............SerSerGluAspArgSerAlaLeuTrpAlaLeuVa 121

633 TACGTTCTATGGGGGAGATTGCCAGCTAACCCTCAATAAGAAATGCACGC 682
    |||||||:::|||||::::|||||||||:::||||||||||||||||||||
121 lThrPheHisGlyGlySerCysGlnLeuAsnLeuAsnLysLysCysThrH 138

683 ATTTGATTGTTCCAGAGCCAAAGGGGGGGAAATACGAATGTGCTTTAAAG 732
    ||||||||||||||||||||||||   |||||||||   |||:::|||
138 isLeuIleValProGluProLysGlyGluLysTyrGluArgAlaValLys 154

733 CGAGCAAGTATTAAAATTGTGACTCCTGACTGGGTTCTGGATTGCGTATC 782
    |||:::||||||||||||||||||||||||||||||||||||||||||||||
155 ArgThrSerIleLysIleValThrProAspTrpValLeuAspCysValSe 171

783 AGAGAAAACCAAAAAGGACGAAGCATTTTATCATCCTCGTCTGATTATTT 832
    |||||||   :::||||||||||||||||||||||||||||||||||||||
171 rGluLysArgArgLysAspGluAlaPheTyrHisProArgLeuIleIleT 188

833 ATGAAGAGGAAGAAGAGGAAGAGGAAGAGGAGGAGGAAGTAGAAAATGAG 882
    |||||||||||||||||||||||||||||          :::||||||
188 yrGluGluGluGluGluGluGluGluGly.........AspAsnGlu 201

883 GAACAAGATTCTCAGAATGAGGGTAGTACAGATGAGAAGTCAAGCCCTGC 932
    |||||||||||||||||||||||||||||||   ||||||||||||   ||
202 GluGlnAspSerGlnAsnGluGlySerThr...GluLysSerSerValAl 217

933 CAGCTCTCAAGAAGGGTCTCCTTCAGGTGACCAGCAGTTTTCACCTAAAT 982
    |||||||      :::|||||:::   :::|||      ||||||||
217 aSerSerAlaValAlaSerProAla...GluGlnProCysSerProLysP 233

983 CCAACACTGAAAAATCTAAAGGGGAATTAATGTTTGATGATTCTTCAGAT 1032
      ::::::|||   ||||||||||||||||||||||||||||||||||||
233 roArgAlaGluValSerLysGlyGluLeuMetPheAspAspSerSerAsp 249

1033 TCATCACCGGAAAAACAGGAGAGAAATTTAAACTGGACCCCGGCCGAAGT 1082
     ||||||||||||||||||||||||:::||||||||||:::||||||||::
250 SerSerProGluLysGlnGluArgSerLeuAsnTrpAlaProAlaGluAl 266

1083 CCCACAGTTAGCTGCAGCAAAACGCAGGCTGCCTCAGGGAAAGGAGCCTG 1132
     :|||   |||   :::|||:::|||||||||||||||||||   ||||
266 aProProLeuAsnThrAlaGlnArgArgLeuProGlnGlyLysGlyProG 283

1133 GGTTGATTAACTTGTGTGCCAATGTCCCACCCGTCCCAGGTAACATTTTG 1182
     ||||||||||||||||||||||||||||||||||||||:::|||||||
283 lyLeuIleAsnLeuCysAlaAsnValProProValProGlyAspIleLeu 299

1183 CCCCCTGAGGTCCGGGGTAATTTAATGGCTGCTGGACAAAACCTCCAAAG 1232
     ||||||::::::|||||||||||||||||   ||||||||||||||::
300 ProProAspMetArgGlyAsnLeuMetAlaProGlyGlnAsnLeuGlnAs 316

1233 TTCTGAAAGATCAGAAATGATAGCTACCTGGAGTCCAGCTGTACGGACAC 1282
     ||||||||||||||||:::::::::|||||||||||||||||||||||||
317 GluArgSerGluIleLeuGlyThrTrpSerProAlaValArgThrL 333

1283 TGAGGAATATTACTAATAATGCTGACATTCAGCAGATGAACCGGCCATCA 1332
     ||||||||||||||||||||||||||||||||||||:::||||||||||||
333 euArgAsnIleThrAsnAsnAlaAspIleGlnGlnIleAsnArgProSer 349

1333 AATGTAGCACATATCTTACAGACTCTTTCAGCACCTACGAAAAATTTAGA 1382
     ||||||||||||||||||||:::||||||||||||||||||||:::||||||
350 AsnValAlaHisIleLeuGlnSerLeuSerAlaProThrLysSerLeuGl 366

1383 ACAGCAGGTGAATCACAGCCAGCAGGGACATACAAATGCCAATGCAGTGC 1432
     ||||||||||   ::::::||||||||||||   |||||||:::||||||||
366 uGlnGlnValAlaArgGlyGlnGlnGlyHisProAsnAlaSerAlaValL 383

1433 TGTTTAGCCAAGTGAAAGTGACTCCAGAGACACACATGCTACAGCAGCAG 1482
     |||||:::|||:::|||   :::||||||||||||:::||||||||:::
383 euPheGlyGlnAlaLysGlyAlaProGluThrHisValLeuGlnGlnHis 399

1483 CAGCAGGCCCAGCAGCAGCAGCAGCAGCACCCGGTTTTACACCTTCAGCC 1532
     :::      ||||||   ||||||||||||||||:::||||||||||||||||
400 HisProProGlnGlnProGlnGlnGlnHisProAlaLeuHisLeuGlnPr 416

1533 CCAGCAGATAATGCAGCTCCAGCAGCAGCAGCAGCAGCAGATCTCTCAG. 1581
     |   ||||||||||||||||||||||||||||||||||||||   :::|||
416 o...GlnIleMetGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnG 432

1582 ..........CAACCTTACCCCCAGCAGCCGCCGCATCCATTTTCA... 1617
                ||||||||||||||||   |||   |||   |||
432 lnGlnGlnProGlnProTyrProGlnProProSerHisGlnPheProGln 448

1617 ..................................................... 1617

449 GlnValHisGlnHisGlnPheSerGlnGlnGlnLeuGlnPheProGlnGl 465

1617 ..................................................... 1617

465 nProLeuHisProGlnGlnGlnLeuHisArgProGlnGlnGlnLeuGlnP 482

1617 ..................................................... 1617

482 roPheGlnGlnGlnHisAlaLeuGlnGlnGlnLeuHisGlnLeuGlnGln 498

1617 ..................................................... 1617

499 GlnGlnLeuGlnHisHisGlnLeuAlaGlnLeuGlnGlnGlnGlnGlnGl 515

1618 ..............CAGCAGCAGCAGCAGCAACAGCAG..... 1641
                  ||||||||||||||||||||||||
515 nGlnHisAsnLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnLeuGlnArgL 532

1641 ..................................................... 1641

532 euGlnGlnGlnGlnGlnMetGlnAsnGlnAlaAlaHisLeuSerGlnAla 548

1641 ..................................................... 1641

549 SerGlnAlaLeuGlnHisGlnValLeuProGlnGlnProLeuGlnLeuSe 565

1642 .........CCACCACCATCGCCTCAGCAGCATCAGCTTTTTGGACATG 1681
             ||||||   :::   ||||||:::||||||||||||||||
565 rLeuGlnProProProGlnGlnGlnGlnGlnGlnLeuPheGlyHisA 582

1682 ATCCAGCAGTGGAGATTCCAGAAGAAGGCTTCTTATTGGGATGTGTGTTT 1731
     ||||||||||||||||||||||||:::||||||||||||||||||||||
582 spProAlaValGluIleProGluGluSerPheLeuLeuGlyCysValPhe 598

1732 GCAATTGCGGATTATCCAGAGCAGATGTCTGATAAGCAACTGCTGGCCAC 1781
     ||||||||||||||||||||||||||||||||||||||||||||||||||
599 AlaIleAlaAspTyrProGluGlnMetSerAspLysGlnLeuLeuAlaTh 615

1782 CTGGAAAAGGATAATCCAGGCACATGGCGGCACTGTTGACCCCACCTTCA 1831
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
615 rTrpLysArgIleIleGlnAlaHisGlyGlyThrValAspProThrPheT 632
```

```
1832 CGAGTCGATGCACGCACCTTCTCTGTGAGAGTCAAGTCAGCAGCGCGTAT 1881
     ||||||||||||||||||||||||||   ||||||||||||||   |||
 632 hrSerArgCysThrHisLeuLeuCysAlaSerGlnValSerSerMetTyr 648

1882 GCACAGGCAATAAGAGAAAGAAAGAGATGTGTTACTGCACACTGGTTAAA 1931
     :::||||||:::||||||||||||||||||||||||||||||||||||||
 649 ThrGlnAlaLeuArgGluArgLysArgCysValThrAlaHisTrpLeuAs 665

1932 CACAGTCTTAAAAAAAAAAAAAAATGGTACCGCCGCACCGAGCCCTTCACT 1981
     ||||||||||||||||||||||:::::::||||||||||||||||||||||
 665 nThrValLeuLysLysLysLysLeuMetProProHisArgAlaLeuHisP 682

1982 TCCCAGTGGCCTTCCCACCAGGAGGAAAGCCATGTTCACAGCATATTATT 2031
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 682 heProValAlaPheProProGlyGlyLysProCysSerGlnHisIleIle 698

2032 TCTGTGACTGGATTTGTTGATAGTGACAGAGATGACCTAAAATTAATGGC 2081
     ||||||||||||||||||||||:::|||||||||||||||||||||||||
 699 SerValThrGlyPheValAspAsnAspArgAspAspLeuLysLeuMetAl 715

2082 TTATTTGGCAGGTGCCAAATATACGGGTTATCTATGCCGCAGCAACACAG 2131
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 715 aTyrLeuAlaGlyAlaLysTyrThrGlyTyrLeuCysArgSerAsnThrV 732

2132 TCCTCATCTGTAAAGAACCAACTGGTTTAAAGTATGAAAAAGCCAAAGAG 2181
     ||||||||||||||||||||:::|||||||||||||||||||||||||||
 732 alLeuIleCysLysGluProSerGlyLeuLysTyrGluLysAlaLysGlu 748

2182 TGGAGGATACCCTGTGTCAACGCCCAGTGGCTTGGCGACATTCTTCTGGG 2231
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 749 TrpArgIleProCysValAsnAlaGlnTrpLeuGlyAspIleLeuLeuGl 765

2232 AAACTTTGAGGCACTGAGGCAGATTCAGTATAGTCGCTACACGGCATTCA 2281
     ||||||||||||||||||||||:::|||||||||||||||||||||||||:
 765 yAsnPheGluAlaLeuArgGlnValGlnTyrSerArgTyrThrAlaPheA 782

2282 GTCTGCAGGATCCATTTGCCCCTACCCAGCATTTAGTTTTAAATCTTTTA 2331
     :::::   ||||||||||:::|||||||   ||||||||||||:::|||||
 782 snMetProAspProPheValProThrProHisLeuValLeuGlyLeuLeu 798

2332 GATGCTTGGAGAGTTCCCTTAAAAGTGTCTGCAGAGTTGTTGATGAGTAT 2381
     ||||||||||||:::|||:::||||||:::|||||||||||||||||:::::
 799 AspAlaTrpArgThrProValLysValThrAlaGluLeuLeuMetGlyVa 815

2382 AAGACTACCTCCCAAACTGAAACAGAATGAAGTAGCTAATGTCCAGCCTT 2431
     :||||||   ||||||||||||   ||||||||||||||:::|||   |
 815 lArgLeu***ProLysLeuLysProAsnGluValAlaAsnIleGlnLeuS 832

2432 CTTCCCAAAGAGCCAGAATTGAAGACGTACCACCTCCCACTAAAAAGCTA 2481
     |||||:::|||||||||||||||||:::|||||||||||||||||||||
 832 erSerLysArgAlaArgIleGluAspLeuProProProThrLysLysLeu 848

2482 ACTCCAGAATTGACCCCTTTTGTGCTTTTCACTGGATTCGAGCCTGTCCA 2531
     ||||||||||||||||:::|||||||||||||||||||||||||||||||
 849 ThrProGluLeuThrProLeuValLeuPheThrGlyPheGluProValGl 865

2532 GGTTCAACAGTATATTAAGAAGCTCTACATTCTTGGTGGAGAGGTTGCGG 2581
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 865 nValGlnGlnTyrIleLysLysLeuTyrIleLeuGlyGlyGluValAlaG 882

2582 AGTCTGCACAGAAGTGCACACACCTCATTGCCAGCAAAGTGACTCGCACC 2631
     ||   ::::::||||||||||||||||||||||||||||||||||||||
 882 luCysThrLysLysCysThrHisLeuIleAlaSerLysValThrArgThr 898

2632 GTGAAGTTCCTGACGGCGATTTCTGTCGTGAAGCACATAGTGACGCCAGA 2681
     ||||||||||||||||||||||||||||||||||||||||||||||||::
 899 ValLysPheLeuThrAlaIleSerValValLysHisIleValThrProAs 915

2682 GTGGCTGGAAGAATGCTTCAGGTGTCAGAAGTTCATTGATGAGCAGAACT 2731
     :||||||||||||||||:::   |||   ||||||||||||||||||||
 915 pTrpLeuGluGluCysPheLysArgGlnThrPheIleAspGluGlnAsnT 932

2732 ACATTCTCCGAGATGCTGAGGCAGAAGTACTTTTCTCTTTCAGCTTGGAA 2781
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 932 yrIleLeuArgAspAlaGluAlaGluValLeuPheSerPheSerLeuGlu 948

2782 GAATCCTTAAAACGGGCACACGTTTCTCCACTCTTTAAGGCAAAATATTT 2831
     ||||||||||||||||||||||||||||:::|||||:::||||||||||
 949 GluSerLeuLysArgAlaHisValSerProPhePheLysThrLysTyrPh 965

2832 TTACATCACACCTGGAATCTGCCCAAGTCTTTCCACTATGAAGGCAATCG 2881
     ||||||||||||||||||||||||||||:::|||||||||||||||||||
 965 eTyrIleThrProGlyIleCysProSerLeuAlaThrMetLysAlaIleV 982

2882 TAGAGTGTGCAGGAGGAAAGGTGTTATCCAAGCAGCCATCTTTCCGGAAG 2931
     ||||||||||||||||||||||||||:::|||||||||||||||||||||
 982 alGluCysAlaGlyGlyLysValLeuAlaLysGlnProSerPheArgLys 998

2932 CTCATGGAGCACAAGCAGAACTCGAGTTTGTCGGAAATAATTTTAATATC 2981
     ||||||||||||||||||||||:::|||||||||||||||||||||||||
 999 LeuMetGluHisLysGlnAsnLysSerLeuSerGluIleIleLeuIleSe 1015

2982 CTGTGAAAATGACCTTCATTTATGCCGAGAATATTTTGCCAGAGGCATAG 3031
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1015 rCysGluAsnAspLeuHisLeuCysArgGluTyrPheAlaArgGlyIleA 1032

3032 ATGTTCACAATGCAGAGTTCGTTCTGACTGGAGTGCTCACTCAAACGCTG 3081
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1032 spValHisAsnAlaGluPheValLeuThrGlyValLeuThrGlnThrLeu 1048

3082 GACTATGAATCATATAAGTTTAAC 3105
     ||||||||||||||||||||||||
1049 AspTyrGluSerTyrLysPheAsn 1056
```

seq_name: sp_human:O15404

seq_documentation_block:

```
ID   O15404      PRELIMINARY;      PRT;   744 AA.
AC   O15404;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CAGF28 (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=97369492; PubMed=9225980;
RA   Margolis R.L., Abraham M.R., Gatchell S.B., Li S.H., Kidwai A.S.,
RA   Breschel T.S., Stine O.C., Callahan C., McInnis M.G., Ross C.A.;
RT   "cDNAs with long CAG trinucleotide repeats from human brain.";
RL   Hum. Genet. 100:114-122(1997).
DR   EMBL; U80735; AAB91434.1; -.
DR   InterPro; IPR001357; BRCT.
DR   Pfam; PF00533; BRCT; 4.
DR   SMART; SM00292; BRCT; 4.
DR   PROSITE; PS50172; BRCT; 2.
FT   NON_TER       1      1
SQ   SEQUENCE   744 AA;  83747 MW;  8FB1D5EC267A70E0 CRC64;
```

alignment_scores:

```
            Quality: 3841.00              Length:    745
              Ratio:  5.163                Gaps:      1
 Percent Similarity:  99.866   Percent Identity: 99.463
```

alignment_block:

```
 US-09-664-641-10 x O15404    ..

 Align seg 1/1  to: O15404  from: 1  to: 744

     874 GAAAATGAGGAACAAGATTCTCAGAATGAGGGTAGTACAGATGAGAAGTC 923
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
   1 GluAsnGluGluGlnAspSerGlnAsnGluGlySerThrAspGluLysSe 17

 924 AAGCCCTGCCAGCTCTCAAGAAGGGTCTCCTTCAGGTGACCAGCAGTTTT 973
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
  17 rSerProAlaSerSerGlnGluGlySerProSerGlyAspGlnGlnPheS 34

 974 CACCTAAATCCAACACTGAAAAATCTAAAGGGGAATTAATGTTTGATGAT 1023
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
  34 erProLysSerAsnThrGluLysSerLysGlyGluLeuMetPheAspAsp 50

1024 TCTTCAGATTCATCACCGGAAAAACAGGAGAGAAATTTAAACTGGACCCC 1073
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
  51 SerSerAspSerSerProGluLysGlnGluArgAsnLeuAsnTrpThrPr 67

1074 GGCCGAAGTCCCACAGTTAGCTGCAGCAAAACGCAGGCTGCCTCAGGGAA 1123
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
  67 oAlaGluValProGlnLeuAlaAlaAlaLysArgArgLeuProGlnGlyL 84

1124 AGGAGCCTGGGTTGATTAACTTGTGTGCCAATGTCCCACCCGTCCCAGGT 1173
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
  84 ysGluProGlyLeuIleAsnLeuCysAlaAsnValProProValProGly 100

1174 AACATTTTGCCCCCTGAGGTCCGGGGTAATTTAATGGCTGCTGGACAAAA 1223
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 101 AsnIleLeuProProGluValArgGlyAsnLeuMetAlaAlaGlyGlnAs 117

1224 CCTCCAAAGTTCTGAAAGATCAGAAATGATAGCTACCTGGAGTCCAGCTG 1273
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 117 nLeuGlnSerSerGluArgSerGluMetIleAlaThrTrpSerProAlaV 134

1274 TACGGACACTGAGGAATATTACTAATAATGCTGACATTCAGCAGATGAAC 1323
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 134 alArgThrLeuArgAsnIleThrAsnAsnAlaAspIleGlnGlnMetAsn 150

1324 CGGCCATCAAATGTAGCACATATCTTACAGACTCTTTCAGCACCTACGAA 1373
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 151 ArgProSerAsnValAlaHisIleLeuGlnThrLeuSerAlaProThrLy 167

1374 AAATTTAGAACAGCAGGTGAATCACAGCCAGCAGGGACATACAAATGCCA 1423
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 167 sAsnLeuGluGlnGlnValAsnHisSerGlnGlnGlyHisThrAsnAlaA 184

1424 ATGCAGTGCTGTTTAGCCAAGTGAAAGTGACTCCAGAGACACACATGCTA 1473
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 184 snAlaValLeuPheSerGlnValLysValThrProGluThrHisMetLeu 200

1474 CAGCAGCAGCAGCAGGCCCAGCAGCAGCAGCAGCACCCGGTTTTACA 1523
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 201 GlnGlnGlnGlnGlnAlaGlnGlnGlnGlnGlnHisProValLeuHi 217

1524 CCTTCAGCCCCAGCAGATAATGCAGCTCCAGCAGCAGCAGCAGCAGCAGA 1573
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 217 sLeuGlnProGlnGlnIleMetGlnLeuGlnGlnGlnGlnGlnGlnGlnI 234

1574 TCTCTCAGCAACCTTACCCCCAGCAGCCGCCGCATCCATTTTCACAGCAG 1623
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 234 leSerGlnGlnProTyrProGlnGlnProProHisProPheSerGlnGln 250

1624 CAGCAGCAGCAACAGCAGCCACCACCATCGCCTCAGCAGCATCAGCTTTT 1673
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 251 GlnGlnGlnGlnGlnProProProSerProGlnGlnHisGlnLeuPh 267

1674 TGGACATGATCCAGCAGTGGAGATTCCAGAAGAAGGCTTCTTATTGGGAT 1723
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 267 eGlyHisAspProAlaValGluIleProGluGluGlyPheLeuLeuGlyC 284

1724 GTGTGTTTGCAATTGCGGATTATCCAGAGCAGATGTCTGATAAGCAACTG 1773
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 284 ysValPheAlaIleAlaAspTyrProGluGlnMetSerAspLysGlnLeu 300

1774 CTGGCCACCTGGAAAAGGATAATCCAGGCACATGGCGGCACTGTTGACCC 1823
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
      LeuAlaThrTrpLysArgIleIleGlnAlaHisGlyGlyThrValAsp.P 317

1824 CACCTTCACGAGTCGATGCACGCACCTTCTCTGTGAGAGTCAAGTCAGCA 1873
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 317 roProSerArgValAspAlaArgThrPheSerValArgValLysSerAla 333

1874 GCGCGTAT.GCACAGGCAATAAGAGAAAGAAAGAGATGTGTTACTGCACA 1922
     ||||||   ||||||||||||||||||||||||||||||||||||||||||
 334 AlaArgIleAlaGlnAlaIleArgGluArgLysArgCysValThrAlaHi 350

1923 CTGGTTAAACACAGTCTTAAAAAAAAAAAAAAAATGGTACCGCCGCACCGAG 1972
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 350 sTrpLeuAsnThrValLeuLysLysLysLysMetValProProHisArgA 367

1973 CCCTTCACTTCCCAGTGGCCTTCCCACCAGGAGGAAAGCCATGTTCACAG 2022
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 367 laLeuHisPheProValAlaPheProProGlyGlyLysProCysSerGln 383

2023 CATATTATTTCTGTGACTGGATTTGTTGATAGTGACAGAGATGACCTAAA 2072
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 384 HisIleIleSerValThrGlyPheValAspSerAspArgAspAspLeuLy 400

2073 ATTAATGGCTTATTTGGCAGGTGCCAAATATACGGGTTATCTATGCCGCA 2122
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 400 sLeuMetAlaTyrLeuAlaGlyAlaLysTyrThrGlyTyrLeuCysArgS 417

2123 GCAACACAGTCCTCATCTGTAAAGAACCAACTGGTTTAAAGTATGAAAAA 2172
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 417 erAsnThrValLeuIleCysLysGluProThrGlyLeuLysTyrGluLys 433

2173 GCCAAAGAGTGGAGGATACCCTGTGTCAACGCCCAGTGGCTTGGCGACAT 2222
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 434 AlaLysGluTrpArgIleProCysValAsnAlaGlnTrpLeuGlyAspIl 450

2223 TCTTCTGGGAAACTTTGAGGCACTGAGGCAGATTCAGTATAGTCGCTACA 2272
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 450 eLeuLeuGlyAsnPheGluAlaLeuArgGlnIleGlnTyrSerArgTyrT 467

2273 CGGCATTCAGTCTGCAGGATCCATTTGCCCCTACCCAGCATTTAGTTTTA 2322
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 467 hrAlaPheSerLeuGlnAspProPheAlaProThrGlnHisLeuValLeu 483

2323 AATCTTTTAGATGCTTGGAGAGTTCCCTTAAAAGTGTCTGCAGAGTTGTT 2372
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 484 AsnLeuLeuAspAlaTrpArgValProLeuLysValSerAlaGluLeuLe 500

2373 GATGAGTATAAGACTACCTCCCAAACTGAAACAGAATGAAGTAGCTAATG 2422
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 500 uMetSerIleArgLeuProProLysLeuLysGlnAsnGluValAlaAsnV 517

2423 TCCAGCCTTCTTCCCAAAGAGCCAGAATTGAAGACGTACCACCTCCCACT 2472
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 517 alGlnProSerSerGlnArgAlaArgIleGluAspValProProProThr 533

2473 AAAAAGCTAACTCCAGAATTGACCCCTTTTGTGCTTTTCACTGGATTCGA 2522
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 534 LysLysLeuThrProGluLeuThrProPheValLeuPheThrGlyPheGl 550

2523 GCCTGTCCAGGTTCAACAGTATATTAAGAAGCTCTACATTCTTGGTGGAG 2572
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 550 uProValGlnValGlnGlnTyrIleLysLysLeuTyrIleLeuGlyGlyG 567

2573 AGGTTGCGGAGTCTGCACAGAAGTGCACACACCTCATTGCCAGCAAAGTG 2622
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 567 luValAlaGluSerAlaGlnLysCysThrHisLeuIleAlaSerLysVal 583

2623 ACTCGCACCGTGAAGTTCCTGACGGCGATTTCTGTCGTGAAGCACATAGT 2672
     ||||||||||:::||||||||||:::||||||||||||||||||||||||
 584 ThrArgThrLeuLysPheLeuAlaAlaIleSerValValLysHisIleVa 600

2673 GACGCCAGAGTGGCTGGAAGAATGCTTCAGGTGTCAGAAGTTCATTGATG 2722
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 600 lThrProGluTrpLeuGluGluCysPheArgCysGlnLysPheIleAspG 617
```

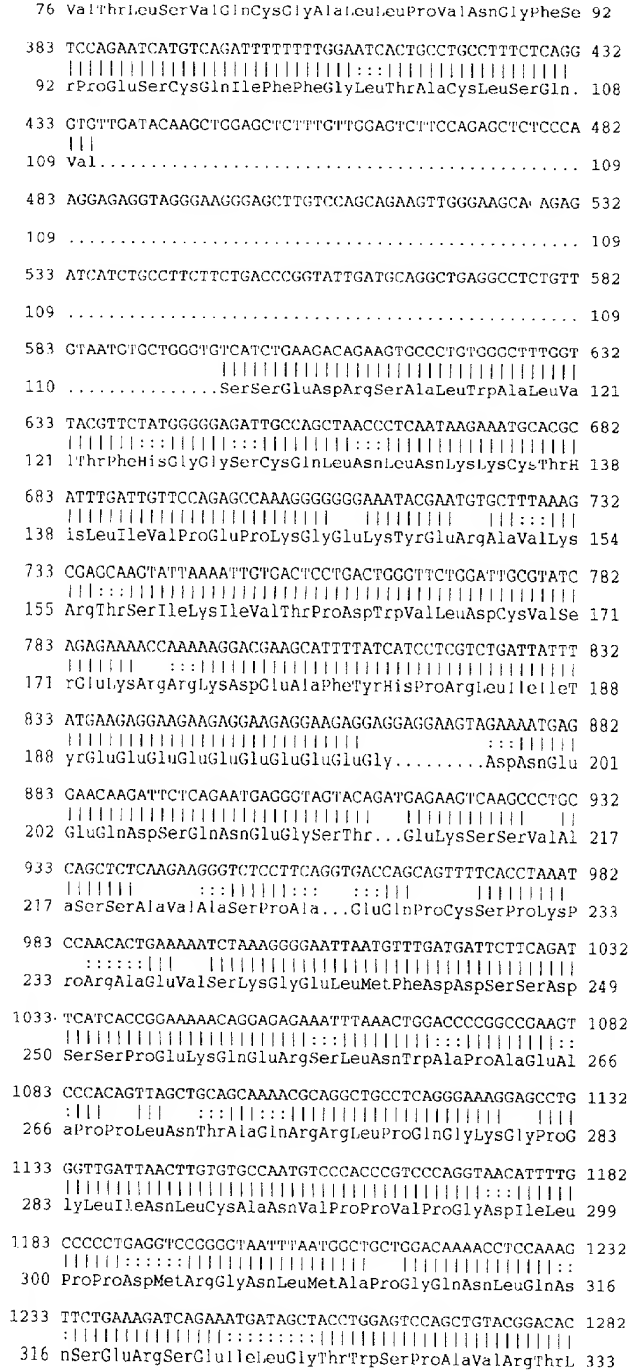
```
2723 AGCAGAACTACATTCTCCGAGATGCTGAGGCAGAAGTACTTTTCTCTTTC 2772
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 617 luGlnAsnTyrIleLeuArgAspAlaGluAlaGluValLeuPheSerPhe 633

2773 AGCTTGGAAGAATCCTTAAAACGGGCACACGTTTCTCCACTCTTTAAGGC 2822
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 634 SerLeuGluGluSerLeuLysArgAlaHisValSerProLeuPheLysAl 650

2823 AAAATATTTTTACATCACACCTGGAATCTGCCCAAGTCTTTCCACTATGA 2872
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 650 aLysTyrPheTyrIleThrProGlyIleCysProSerLeuSerThrMetL 667

2873 AGGCAATCGTAGAGTGTGCAGGAGGAAAGGTGTTATCCAAGCAGCCATCT 2922
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 667 ysAlaIleValGluCysAlaGlyGlyLysValLeuSerLysGlnProSer 683

2923 TTCCGGAAGCTCATGGAGCACAAGCAGAACTCGAGTTTGTCGGAAATAAT 2972
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 684 PheArgLysLeuMetGluHisLysGlnAsnSerSerLeuSerGluIleIl 700

2973 TTTAATATCCTGTGAAAATGACCTTCATTTATGCCGAGAATATTTTGCCA 3022
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 700 eLeuIleSerCysGluAsnAspLeuHisLeuCysArgGluTyrPheAlaA 717

3023 GAGGCATAGATGTTCACAATGCAGAGTTCGTTCTGACTGGAGTGCTCACT 3072
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 717 rgGlyIleAspValHisAsnAlaGluPheValLeuThrGlyValLeuThr 733

3073 CAAACGCTGGACTATGAATCATATAAGTTTAAC 3105
     |||||||||||||||||||||||||||||||||
 734 GlnThrLeuAspTyrGluSerTyrLysPheAsn 744
```

seq_name: sp_vertebrate:Q90WJ3

```
seq_documentation_block:
ID   Q90WJ3      PRELIMINARY;      PRT;  1256 AA.
AC   Q90WJ3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SWIFT.
GN   K14.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE=21260044; PubMed=11359898;
RA   Shimizu K., Bourillot P.Y., Nielsen S.J., Zorn A.M., Gurdon J.B.;
RT   "Swift Is a Novel BRCT Domain Coactivator of Smad2 in Transforming
RT   Growth Factor beta Signaling.";
RL   Mol. Cell. Biol. 21:3901-3912(2001).
DR   EMBL; AF172855; AAK55123.1; -.
SQ   SEQUENCE   1256 AA;  144187 MW;  E7795B12C1A42DE1 CRC64;
```

```
alignment_scores:
            Quality: 3425.50            Length:   1332
              Ratio:  4.112              Gaps:     20
 Percent Similarity:  62.538   Percent Identity: 53.829

alignment_block:
 US-09-664-641-10 x Q90WJ3   ..

 Align seg 1/1  to: Q90WJ3  from: 1  to: 1256

165 CCTTCAAATCACCTCTAGGCACAGGTTATTCAGCTTCTCAAGGCTGGAAA 214
    |||   :::   |||      :::|||:::|||:::|||||||||||||||
 20 ProValGlyAspLeuAspProLysValValGlnMetLeuLysAlaGlyLy 36

215 AGCGAAGGAAGTTTCCTACAATGCACTAGCCTCACACATAATCTCAGAGG 264
    |||||||||||||||||||||||||||||||:::|||||||||:::||||
 36 sAlaLysGluValSerTyrAsnAlaLeuAlaThrHisIleIleAlaGluA 53

265 ATGGGGACAATCCAGAGGTGGGAGAAGCTCGGGAAGTCTTTGACTTACCT 314
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
 53 spGlyAspAsnProGluValGlyGluAlaArgGluValPheAspLeuPro 69

315 GTTGTAAAGCCTTCTTGGGTGATTCTGTCCGTTCAGTGTGGAACTCTTCT 364
    ||||||||||||||||||||||||||||||||:::||||||:::   ||
 70 ValValLysProSerTrpValIleLeuSerValArgCysGlyValArgLe 86

365 GCCAGTAAATGGTTTTTCTCCAGAATCATGTCAGATTTTTTTTGGAATCA 414
    ||||   ||||||||||||||||||||   ||||||||||||||||:::|
 86 uProGluAsnGlyPheSerProGluSerGlyGlnIlePhePheGlyValT 103

415 CTGCCTGCCTTTCTCAGGGTGTTGATACAAGCTGGAGCTCTTTGTTGGAG 464
    |||||||||||||||| |||
103 hrAlaCysLeuSerGln.Val..................................... 109

465 TCTTCCAGAGCTCTCCCAAGGAGAGGTAGGGAAGGGAGCTTGTCCAGCAG 514

109 ..... 109

515 AAGTTGGGAAGCACAGAGATCATCTGCCTTCTTCTGACCCGGTATTGATG 564

109 ..... 109

565 CAGGCTGAGGCCTCTGTTGTAATGTGCTGGGTGTCATCTGAAGACAGAAG 614
                                        |||   :::|||||||::
110 ..............................SerProAspAspArgAs 115

615 TGCCCTGTGGGCTTTGGTTACGTTCTATGGGGGAGATTGCCAGCTAACCC 664
    ::::|||||||||||||:::|||||||||||||||||||||||||||:::|
115 nSerLeuTrpAlaLeuThrThrPheTyrGlyGlyAspCysGlnLeuSerL 132

665 TCAATAAGAAATGCACGCATTTGATTGTTCCAGAGCCAAAGGGGGGGGAAA 714
    ||||||||||||||||||||||||||||||||||||||||||||||:::|||
132 euAsnLysLysCysThrHisLeuIleValProGluProLysGlyAsnLys 148

715 TACGAATGTGCTTTAAAGCGAGCAAGTATTAAAATTGTGACTCCTGACTG 764
    ||||||   |||::::::|||:::|||||||||||||||||||||||||||
149 TyrGluTyrAlaPheGlnArgGlySerIleLysIleValThrProAspTr 165

765 GGTTCTGGATTGCGTATCAGAGAAAACCAAAAAGGACGAAGCATTTTATC 814
    ||||||||||   |||||||||||||||||||||||||||||:::||||
165 pValLeuAspSerValSerGluLysThrLysLysAspGluAlaLeuTyrH 182

815 ATCCTCGTCTGATTATTTATGAAGAGGAAGAAGAGGAAGAGGAAGAGGAG 864
    ||||||||||||||:::|||       ||||||||||||||||||||||
182 isProArgLeuIleValTyr......GluGluGluGluGluGluGluGlu 196

865 GAGGAAGTAGAAAATGAGGAACAAGATTCTCAGAATGAGGGTAGTACAGA 914
    |||||||:::|||||||||||||||:::|||||||||   ::::::   ||
197 GluGluAlaGluAsnGluGluGlnGluSerGlnAsnGlySerAspGluAs 213

915 TGAG.....AAGTCAAGCCCTGCCAGCTCTCAAGAAGGGTCTCCTT 955
    |:::         ||||||||||||:::::::|||:::::::|||||||||
213 pLysTyrSerSerLysSerSerProValThrSerArgAspGlySerProV 230

956 CAGGTGACCAGCAGTTTTCACCTAAA...TCCAACACTGAAAAATCTAAA 1002
        :::::::::   |||||||||   |||:::::::::|||   |||
230 alHisAsnArgGluSerSerProLysArgSerSerSerAspLysIleLys 246

1003 GGGGAATTAATGTTTGATGATTCTTCAGATTCATCACCGGAAAAACAGGA 1052
     :::||||||||||||||||||||||||||||||||||||||||||||||||
247 SerGluLeuMetPheAspAspSerSerAspSerSerProGluLysGlnGl 263

1053 GAGAAATTTAAACTGGACCCCGGCCGAAGTCCCACAGTTAGCTGCAGCAA 1102
     ||||||||||||||||||||||||||||||||||||||:::::::::||||
263 uArgAsnLeuAsnTrpThrProAlaGluValProGlnMetSerThrAlaL 280

1103 AACGCAGGCTGCCTCAGGGAAAGGAGCCTGGGTTGATTAACTTGTGTGCC 1152
```

```
OM of: US-09-664-641-10 to: SPTREMBL_19:*  out_format : pfs

Date: Jun 4, 2002  5:49 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_n2p.model -DEV=x1p
 -Q=/cgn2_1/USPTO_spool/US09664641/runat_04062002_110933_9250/app_query.fasta_1.3697
 -DB=SPTREMBL_19 -QFMT=fastan -SUFFIX=rspt -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09664641_@CGN1_1_312 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-664-641-10
Query length: 3580
Database: SPTREMBL_19:*
Database sequences: 562222
Database length: 172994929
Search time (sec): 230.300000



score_list:
Sequence              Strd Orig    %Score   EScore Len  ! Documentation  ..
sp_rodent:Q9Z0W6      + 4059.50 5875.12  0.0       1056  ! Q9z0w6 mus musculus (mouse). p
sp_human:O15404       + 3841.00 5561.10  0.0       744  ! O15404 homo sapiens (human). ca
sp_vertebrate:Q90WJ3 + 3425.50 4951.40  1.5e-268  1256  ! Q90wj3 xenopus laevis (africa
sp_human:Q96HP2       + 2039.00 2947.67  1.1e-156  391  ! Q96hp2 homo sapiens (human). un
sp_invertebrate:Q9VUB6 +  852.00 1205.56  4.6e-60   1798  ! Q9vub6 drosophila melanogast
sp_invertebrate:Q94046 +  524.00  734.07  3.4e-33   1074  ! Q94046 caenorhabditis elegan
sp_invertebrate:Q9U370 +  524.00  734.05  1.4e-33   1076  ! Q9u370 caenorhabditis elegan
sp_human:Q14676       +  342.50  463.12  8.9e-19   2089  ! Q14676 homo sapiens (human). k
sp_human:Q96QC2       +  340.50  460.21  1.3e-18   2090  ! Q96qc2 homo sapiens (human). k
sp_plant:Q9LVF7       +  316.50  432.68  8.9e-17   1041  ! Q9lvf7 arabidopsis thaliana (m
sp_rodent:Q91VX3      +  283.50  382.38  4.5e-14   1296  ! Q91vx3 mus musculus (mouse). s
sp_plant:Q9ZR18       +  265.50  361.78  1.1e-12   765  ! Q9zr18 arabidopsis thaliana (mo
sp_human:Q9UEB9       +  263.50  352.22  2.0e-12   1435  ! Q9ueb9 homo sapiens (human). d
sp_human:Q92547       +  263.50  351.40  2.0e-12   1550  ! Q92547 homo sapiens (human). k
sp_fungi:Q9P5J5       +  229.00  307.74  1.0e-09   838  ! Q9p5j5 neurospora crassa. conse
sp_invertebrate:Q9VFR7 +  225.50  303.46  1.9e-09   7[illegible]  ! Q9vfr7 drosophila melanogaste
sp_invertebrate:Q24163 +  220.50  291.76  5.5e-09   1180  ! Q24163 drosophila melanogast
sp_invertebrate:Q9VED3 +  220.50  291.71  5.6e-09   1186  ! Q9ved3 drosophila melanogast
sp_invertebrate:O44011 +  219.50  288.08  7.2e-09   1457  ! O44011 dictyostelium discoid
sp_invertebrate:Q9W4E8 +  216.00  293.94  9.6e-09   517  ! Q9w4e8 drosophila melanogaste
sp_plant:Q93VQ3       +  215.50  283.70  1.4e-08   1272  ! Q93vq3 oryza sativa (rice). hy
sp_invertebrate:O76853 +  214.00  293.23  1.3e-08   420  ! O76853 dictyostelium discoide
sp_fungi:P79025       +  214.00  291.75  1.4e-08   483  ! P79025 emericella nidulans (asp
sp_virus:Q98148       +  214.00  282.47  1.9e-08   1162  ! Q98148 kaposi's sarcoma-associ
sp_invertebrate:Q9VP17 +  213.50  274.60  2.6e-08   2285  ! Q9vp17 drosophila melanogast
sp_vertebrate:Q90Y12 +  212.50  275.54  2.9e-08   1822  ! Q90yi2 fugu rubripes (japanese
sp_invertebrate:Q9Y0C9 +  212.00  283.02  2.4e-08   838  ! Q9y0c9 dictyostelium discoide
sp_invertebrate:O18346 +  210.50  278.87  3.4e-08   1010  ! O18346 drosophila hydei (fru
sp_virus:Q9QR71       +  210.00  276.96  3.9e-08   1129  ! Q9qr71 kaposi's sarcoma-associ
sp_invertebrate:Q9VMS2 +  208.50  266.64  6.7e-08   2439  ! Q9vms2 drosophila melanogast
sp_invertebrate:Q95ZG5 +  207.00  262.29  9.6e-08   2994  ! Q95zg5 dictyostelium discoid
sp_invertebrate:Q9V7C5 +  206.50  273.98  6.9e-08   925  ! Q9v7c5 drosophila melanogaste
sp_invertebrate:Q95YM8 +  206.50  268.20  8.4e-08   1598  ! Q95ym8 apis mellifera (honey
sp_invertebrate:Q17641 -  205.50  288.88  4.8e-08   197  ! Q17641 caenorhabditis elegans
sp_invertebrate:Q9VN59 +  205.00  271.70  9.2e-08   934  ! Q9vn59 drosophila melanogaste
sp_invertebrate:Q94485 +  204.50  274.64  8.9e-08   660  ! Q94485 dictyostelium discoide
sp_invertebrate:Q9N8T2 -  203.00  255.01  2.1e-07   3439  ! Q9n8t2 trypanosoma brucei. p
sp_human:O15410       +  202.50  271.86  1.3e-07   652  ! O15410 homo sapiens (human). ca
sp_human:O75557       +  202.50  259.89  1.9e-07   2023  ! O75557 homo sapiens (human). d
sp_human:Q9UND7       +  202.50  259.89  1.9e-07   2023  ! Q9und7 homo sapiens (human). d
sp_virus:Q9DUN0       +  201.00  265.41  2.0e-07   976  ! Q9dun0 kaposi's sarcoma-asso
sp_rodent:Q64507      -  200.00  279.24  1.4e-07   230  ! Q64507 mus musculus (mouse).
sp_human:Q9BT44       +  200.00  268.27  2.0e-07   649  ! Q9bt44 homo sapiens (human).
sp_virus:O40947       +  200.00  262.80  2.5e-07   1089  ! O[illegible]947 kaposi's sarcoma-ass
sp_invertebrate:Q9U9S7 +  200.00  255.74  3.1e-07   2123  ! Q9u9s7 dictyostelium disc

seq_name: sp_rodent:Q9Z0W6

seq_documentation_block:
ID   Q9Z0W6      PRELIMINARY;      PRT;  1056 AA.
AC   Q9Z0W6;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PAX TRANSCRIPTION ACTIVATION DOMAIN INTERACTING PROTEIN PTIP.
GN   PAXIP1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB;
RX   MEDLINE=20368635; PubMed=10908331;
RA   Lechner M.S., Levitan I., Dressler G.R.;
RT   "PTIP, a novel BRCT domain-containing protein interacts with Pax2 and
RT   is associated with active chromatin.";
RL   Nucleic Acids Res. 28:2741-2751(2000).
DR   EMBL; AF104261; AAD17923.1; -.
DR   MGD; MGI:1890430; Paxip1.
DR   InterPro; IPR001357; BRCT.
DR   Pfam; PF00533; BRCT; 6.
DR   SMART; SM00292; BRCT; 5.
DR   PROSITE; PS50172; BRCT; 5.
SQ   SEQUENCE   1056 AA;  119331 MW;  99CAF4EF4EE53ED6 CRC64;


alignment_scores:
            Quality: 4059.50              Length:   1142
              Ratio:  4.572                 Gaps:     11
 Percent Similarity:  77.758   Percent Identity: 70.928

alignment_block:
 US-09-664-641-10 x Q9Z0W6   ..

 Align seg 1/1  to: Q9Z0W6  from: 1  to: 1056

      33 CCGGAATTCCCGGGTCGACGATTTCGTCGAGTGCAAGACGGGCGGGCTGC 82
         ||||||   |||   :::   ||||||:::|||:::
       6 ProGluValProGluGluLeuPheArgGluValLys..... 17

      83 TGAGGGCGCCCCGCGGGTCTGGTCTGGGCCGGGGAACTCCGGGACGGCGC 132

      17 ..................................................... 17

     133 CGCAGCAGGGCTACGGATCTCTCGGCCCCTGCCCTTCAAATCACCTCTAG 182
                :::   |||::::::   |||
      18 ..TyrTyrAlaValGlyAspIleAspPro..... 26

     183 GCACAGGTTATTCAGCTTCTCAAGGCTGGAAAAGCGAAGGAAGTTTCCTA 232
           ||||||||||||||||||||||||||||||||||||||||||||||||||
      27 ...GlnValIleGlnLeuLeuLysAlaGlyLysAlaLysGluValSerTy 42

     233 CAATGCACTAGCCTCACACATAATCTCAGAGGATGGGGACAATCCAGAGG 282
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      42 rAsnAlaLeuAlaSerHisIleIleSerGluAspGlyAspAsnProGluV 59

     283 TGGGAGAAGCTCGGGAAGTCTTTGACTTACCTGTTGTAAAGCCTTCTTGG 332
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      59 alGlyGluAlaArgGluValPheAspLeuProValValLysProSerTrp 75

     333 GTGATTCTGT[illegible] GTTCAGTGTGGAACTCTTCTGCCAGTAAATGGTTTTTC 382
         |||   ||||||||||||||||||:::|||||||||||||||||||||||||
```


76 ValThrLeuSerValGlnCysGlyAlaLeuLeuProValAsnGlyPheSe 92
383 TCCAGAATCATGTCAGATTTTTTTTGGAATCACTGCCTGCCTTTCTCAGG 432
92 rProGluSerCysGlnIlePhePheGlyLeuThrAlaCysLeuSerGln. 108
433 GTGTTGATACAAGCTCGAGCTCTTTGTTGGAGTCTTCCAGAGCTCTCCCA 482
109 Val..... 109
483 AGGAGAGGTAGGGAAGGGAGCTTGTCCAGCAGAAGTTGGGAAGCA AGAG 532
109 109
533 ATCATCTGCCTTCTTCTGACCCGGTATTGATGCAGGCTGAGGCCTCTGTT 582
109 109
583 GTAATCTGCTGGGTGTCATCTGAAGACAGAAGTGCCCCTGTGGGCTTTGGT 632
110 .....SerSerGluAspArgSerAlaLeuTrpAlaLeuVa 121
633 TACGTTCTATGGGGGAGATTGCCAGCTAACCCTCAATAAGAAATGCACGC 682
121 lThrPheHisGlyGlySerCysGlnLeuAsnLeuAsnLysLysCysThrH 138
683 ATTTGATTGTTCCAGAGCCAAAGGGGGGGAAATACGAATGTGCTTTAAAG 732
138 isLeuIleValProGluProLysGlyGluLysTyrGluArgAlaValLys 154
733 CGAGCAAGTATTAAAATTGTGACTCCTGACTGGGTTCTGGATTGCGTATC 782
155 ArgThrSerIleLysIleValThrProAspTrpValLeuAspCysValSe 171
783 AGAGAAAACCAAAAAGGACGAAGCATTTTATCATCCTCGTCTGATTATTT 832
171 rGluLysArgArgLysAspGluAlaPheTyrHisProArgLeuIleIleT 188
833 ATGAAGAGGAAGAAGAGGAAGAGGAAGAGGAGGAGGAAGTAGAAAATGAG 882
188 yrGluGluGluGluGluGluGluGluGly.....AspAsnGlu 201
883 GAACAAGATTCTCAGAATGAGGGTAGTACAGATGAGAAGTCAAGCCCTGC 932
202 GluGlnAspSerGlnAsnGluGlySerThr...GluLysSerSerValAl 217
933 CAGCTCTCAAGAAGGGTCTCCTTCAGGTGACCAGCAGTTTTCACCTAAAT 982
217 aSerSerAlaValAlaSerProAla...GluGlnProCysSerProLysP 233
983 CCAACACTGAAAAATCTAAAGGGGAATTAATGTTTGATGATTCTTCAGAT 1032
233 roArgAlaGluValSerLysGlyGluLeuMetPheAspAspSerSerAsp 249
1033 TCATCACCGGAAAAACAGGAGAGAAATTTAAACTGGACCCCGGCCCAAGT 1082
250 SerSerProGluLysGlnGluArgSerLeuAsnTrpAlaProAlaGluAl 266
1083 CCCACAGTTAGCTGCAGCAAAACGCAGGCTGCCTCAGGGAAAGGAGCCTG 1132
266 aProProLeuAsnThrAlaGlnArgArgLeuProGlnGlyLysGlyProG 283
1133 GGTTGATTAACTTGTGTGCCAATGTCCCACCCGTCCCAGGTAACATTTTG 1182
283 lyLeuIleAsnLeuCysAlaAsnValProProValProGlyAspIleLeu 299
1183 CCCCCTGAGGTCCGGGGTAATTTAATGGCTGCTGGACAAAACCTCCAAAG 1232
300 ProProAspMetArgGlyAsnLeuMetAlaProGlyGlnAsnLeuGlnAs 316
1233 TTCTGAAAGATCAGAAATGATAGCTACCTGGAGTCCAGCTGTACGGACAC 1282
316 nSerGluArgSerGluIleLeuGlyThrTrpSerProAlaValArgThrL 333
1283 TGAGGAATATTACTAATAATGCTGACATTCAGCAGATGAACCGGCCATCA 1332
333 euArgAsnIleThrAsnAsnAlaAspIleGlnGlnIleAsnArgProSer 349
1333 AATGTAGCACATATCTTACAGACTCTTTCAGCACCTACGAAAAATTTAGA 1382
350 AsnValAlaHisIleLeuGlnSerLeuSerAlaProThrLysSerLeuGl 366
1383 ACAGCAGGTGAATCACAGCCAGCAGGGACATACAAATGCCAATGCAGTGC 1432
366 uGlnGlnValAlaArgGlyGlnGlnGlyHisProAsnAlaSerAlaValL 383
1433 TGTTTAGCCAAGTGAAAGTGACTCCAGAGACACACATGCTACAGCAGCAG 1482
383 euPheGlyGlnAlaLysGlyAlaProGluThrHisValLeuGlnGlnHis 399
1483 CAGCAGGCCCAGCAGCAGCAGCAGCAGCACCCGGTTTTACACCTTCAGCC 1532
400 HisProProGlnGlnProGlnGlnGlnHisProAlaLeuHisLeuGlnPr 416
1533 CCAGCAGATAATGCAGCTCCAGCAGCAGCAGCAGCAGCAGATCTCTCAG. 1581
416 o...GlnIleMetGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnG 432
1582 .....CAACCTTACCCCCAGCAGCCGCCGCATCCATTTTCA... 1617
432 lnGlnGlnProGlnProTyrProGlnProProSerHisGlnPheProGln 448
1617 1617
449 GlnValHisGlnHisGlnPheSerGlnGlnGlnLeuGlnPheProGlnGl 465
1617 1617
465 nProLeuHisProGlnGlnGlnLeuHisArgProGlnGlnGlnLeuGlnP 482
1617 1617
482 roPheGlnGlnGlnHisAlaLeuGlnGlnGlnLeuHisGlnLeuGlnGln 498
1617 1617
499 GlnGlnLeuGlnHisHisGlnLeuAlaGlnLeuGlnGlnGlnGlnGlnGl 515
1618 .....CAGCAGCAGCAGCAGCAACAGCAG..... 1641
515 nGlnHisAsnLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnLeuGlnArgL 532
1641 1641
532 euGlnGlnGlnGlnGlnMetGlnAsnGlnAlaAlaHisLeuSerGlnAla 548
1641 1641
549 SerGlnAlaLeuGlnHisGlnValLeuProGlnGlnProLeuGlnLeuSe 565
1642 ..... .CCACCACCATCGCCTCAGCAGCATCAGCTTTTTGGACATG 1681
565 rLeuGlnProProProGlnGlnGlnGlnGlnGlnGlnLeuPheGlyHisA 582
1682 ATCCAGCAGTGGAGATTCCAGAAGAAGGCTTCTTATTGGGATGTGTGTTT 1731
582 spProAlaValGluIleProGluGluSerPheLeuLeuGlyCysValPhe 598
1732 GCAATTGCGGATTATCCAGAGCAGATGTCTGATAAGCAACTGCTGGCCAC 1781
599 AlaIleAlaAspTyrProGluGlnMetSerAspLysGlnLeuLeuAlaTh 615
1782 CTGGAAAAGGATAATCCAGGCACATGGCGGCACTGTTGACCCCACCTTCA 1831
615 rTrpLysArgIleIleGlnAlaHisGlyGlyThrValAspProThrPheT 632

```
1832 CGAGTCCATGCCACCGCACCTTCTCTGTGAGAGTCAAGTCAGCAGCGCGTAT 1881
     ||||||||||||||||||||||||||   ||||||||||||||||   |||
 632 hrSerArgCysThrHisLeuLeuCysAlaSerGlnValSerSerMetTyr 648

1882 GCACAGGCAATAAGAGAAAGAAAGAGATGTGTTACTGCACACTGGTTAAA 1931
     :::|||||||:::|||||||||||||||||||||||||||||||||||||
 649 ThrGlnAlaLeuArgGluArgLysArgCysValThrAlaHisTrpLeuAs 665

1932 CACAGTCTTAAAAAAAAAAAAAAATGGTACCGCCGCACCGAGCCCTTCACT 1981
     ||||||||||||||||||||||:::::|||||||||||||||||||||||
 665 nThrValLeuLysLysLysLysLeuMetProProHisArgAlaLeuHisP 682

1982 TCCCAGTGGCCTTCCCACCAGGAGGAAAGCCATGTTCACAGCATATTATT 2031
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 682 heProValAlaPheProProGlyGlyLysProCysSerGlnHisIleIle 698

2032 TCTGTGACTGGATTTGTTGATAGTGACAGAGATGACCTAAAATTAATGGC 2081
     ||||||||||||||||||||||:::|||||||||||||||||||||||||
 699 SerValThrGlyPheValAspAsnAspArgAspAspLeuLysLeuMetAl 715

2082 TTATTTGGCAGGTGCCAAATATACGGGTTATCTATGCCGCAGCAACACAG 2131
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 715 aTyrLeuAlaGlyAlaLysTyrThrGlyTyrLeuCysArgSerAsnThrV 732

2132 TCCTCATCTGTAAAGAACCAACTGGTTTAAAGTATGAAAAAGCCAAAGAG 2181
     ||||||||||||||||||||:::|||||||||||||||||||||||||||
 732 alLeuIleCysLysGluProSerGlyLeuLysTyrGluLysAlaLysGlu 748

2182 TGGAGGATACCCTGTGTCAACGCCCAGTGGCTTGGCGACATTCTTCTGGG 2231
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 749 TrpArgIleProCysValAsnAlaGlnTrpLeuGlyAspIleLeuLeuGl 765

2232 AAACTTTGAGGCACTGAGGCAGATTCAGTATAGTCGCTACACGGCATTCA 2281
     ||||||||||||||||||||||:::|||||||||||||||||||||||:
 765 yAsnPheGluAlaLeuArgGlnValGlnTyrSerArgTyrThrAlaPheA 782

2282 GTCTGCAGGATCCATTTGCCCCTACCCAGCATTTAGTTTTAAATCTTTTA 2331
     :::::   |||||||||:::|||||||   |||||||||||||:::||||||
 782 snMetProAspProPheValProThrProHisLeuValLeuGlyLeuLeu 798

2332 GATGCTTGGAGAGTTCCCTTAAAAGTGTCTGCAGAGTTGTTGATGAGTAT 2381
     ||||||||||||:::|||:::|||||||:::|||||||||||||||:::::
 799 AspAlaTrpArgThrProValLysValThrAlaGluLeuLeuMetGlyVa 815

2382 AAGACTACCTCCCAAACTGAAACAGAATGAAGTAGCTAATGTCCAGCCTT 2431
     :||||||   ||||||||||||   ||||||||||||||:::|||   |
 815 lArgLeu***ProLysLeuLysProAsnGluValAlaAsnIleGlnLeuS 832

2432 CTTCCCAAAGAGCCAGAATTGAAGACGTACCACCTCCCACTAAAAAGCTA 2481
     |||||:::|||||||||||||||||||:::|||||||||||||||||||
 832 erSerLysArgAlaArgIleGluAspLeuProProProThrLysLysLeu 848

2482 ACTCCAGAATTGACCCCTTTTGTGCTTTTCACTGGATTCGAGCCTGTCCA 2531
     ||||||||||||||||:::|||||||||||||||||||||||||||||||
 849 ThrProGluLeuThrProLeuValLeuPheThrGlyPheGluProValGl 865

2532 GGTTCAACAGTATATTAAGAAGCTCTACATTCTTGGTGGAGAGGTTGCGG 2581
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 865 nValGlnGlnTyrIleLysLysLeuTyrIleLeuGlyGlyGluValAlaG 882

2582 AGTCTGCACAGAAGTGCACACACCTCATTGCCAGCAAAGTGACTCGCACC 2631
     ||   ::::::|||||||||||||||||||||||||||||||||||||||
 882 luCysThrLysLysCysThrHisLeuIleAlaSerLysValThrArgThr 898

2632 GTGAAGTTCCTGACGGCGATTTCTGTCGTGAAGCACATAGTGACGCCAGA 2681
     ||||||||||||||||||||||||||||||||||||||||||||||||::
 899 ValLysPheLeuThrAlaIleSerValValLysHisIleValThrProAs 915

2682 GTGGCTGGAAGAATGCTTCAGGTGTCAGAAGTTCATTGATGAGCAGAACT 2731
     :||||||||||||||||:::   |||   ||||||||||||||||||||
 915 pTrpLeuGluGluCysPheLysArgGlnThrPheIleAspGluGlnAsnT 932

2732 ACATTCTCCGAGATGCTGAGGCAGAAGTACTTTTCTCTTTCAGCTTGGAA 2781
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 932 yrIleLeuArgAspAlaGluAlaGluValLeuPheSerPheSerLeuGlu 948

2782 GAATCCTTAAAACGGGCACACGTTTCTCCACTCTTTAAGGCAAAATATTT 2831
     ||||||||||||||||||||||||||||:::|||||||:::||||||||
 949 GluSerLeuLysArgAlaHisValSerProPheLeuLysThrLysTyrPh 965

2832 TTACATCACACCTGGAATCTGCCCAAGTCTTTCCACTATGAAGGCAATCG 2881
     ||||||||||||||||||||||||||||||:::|||||||||||||||||
 965 eTyrIleThrProGlyIleCysProSerLeuAlaThrMetLysAlaIleV 982

2882 TAGAGTGTGCAGGAGGAAAGGTGTTATCCAAGCAGCCATCTTTCCGGAAG 2931
     ||||||||||||||||||||||||||:::|||||||||||||||||||||
 982 alGluCysAlaGlyGlyLysValLeuAlaLysGlnProSerPheArgLys 998

2932 CTCATGGAGCACAAGCAGAACTCGAGTTTGTCGGAAATAATTTTTAATATC 2981
     ||||||||||||||||||||||:::|||||||||||||||||||||||||
 999 LeuMetGluHisLysGlnAsnLysSerLeuSerGluIleIleLeuIleSe 1015

2982 CTGTGAAAATGACCTTCATTTATGCCGAGAATATTTTGCCAGAGGCATAG 3031
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1015 rCysGluAsnAspLeuHisLeuCysArgGluTyrPheAlaArgGlyIleA 1032

3032 ATGTTCACAATGCAGAGTTCGTTCTGACTGGAGTGCTCACTCAAACGCTG 3081
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1032 spValHisAsnAlaGluPheValLeuThrGlyValLeuThrGlnThrLeu 1048

3082 GACTATGAATCATATAAGTTTAAC 3105
     ||||||||||||||||||||||||
1049 AspTyrGluSerTyrLysPheAsn 1056
```

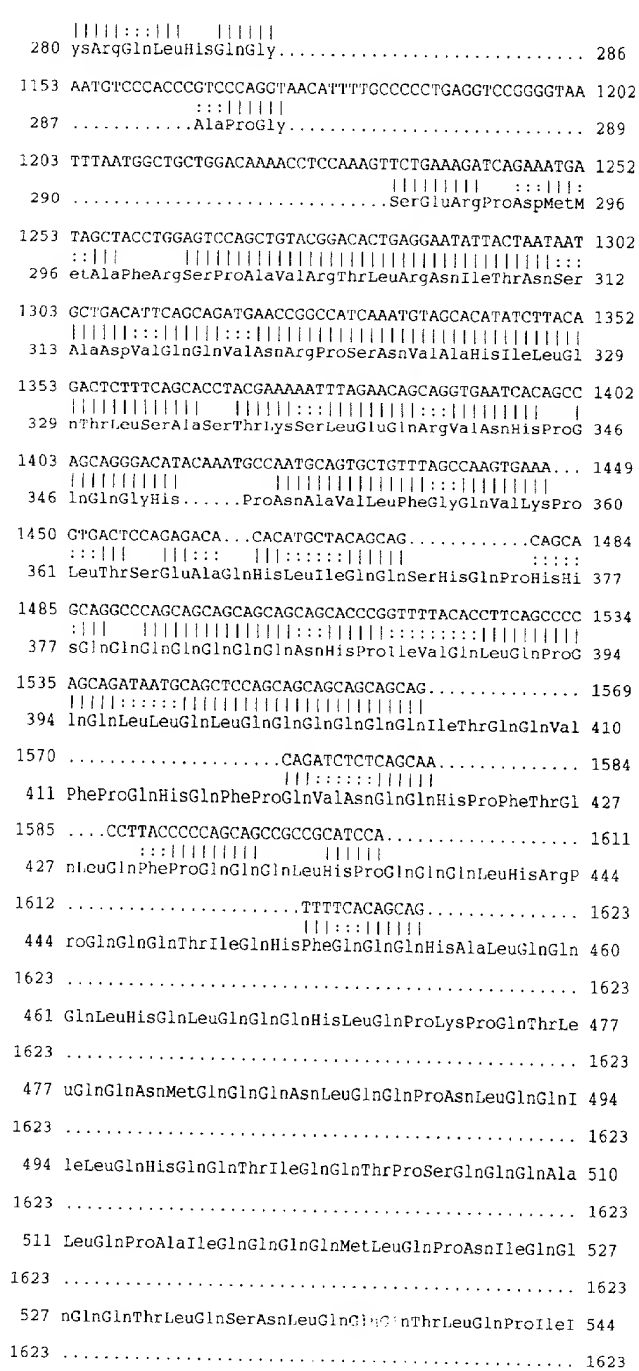
seq_name: sp_human:O15404

```
seq_documentation_block:
ID   O15404      PRELIMINARY;      PRT;   744 AA.
AC   O15404;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CAGF28 (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=97369492; PubMed=9225980;
RA   Margolis R.L., Abraham M.R., Gatchell S.B., Li S.H., Kidwai A.S.,
RA   Breschel T.S., Stine O.C., Callahan C., McInnis M.G., Ross C.A.;
RT   "cDNAs with long CAG trinucleotide repeats from human brain.";
RL   Hum. Genet. 100:114-122(1997).
DR   EMBL; U80735; AAB91434.1; -.
DR   InterPro; IPR001357; BRCT.
DR   Pfam; PF00533; BRCT; 4.
DR   SMART; SM00292; BRCT; 4.
DR   PROSITE; PS50172; BRCT; 2.
FT   NON_TER      1      1
SQ   SEQUENCE   744 AA;  83747 MW;  8FB1D5EC267A70E0 CRC64;
```

```
alignment_scores:
           Quality: 3841.00              Length:    745
             Ratio:  5.163                Gaps:      1
 Percent Similarity:  99.866   Percent Identity: 99.463

alignment_block:
 US-09-664-641-10 x O15404    ..

 Align seg 1/1  to: O15404  from: 1  to: 744

 874 GAAAATGAGGAACAAGATTCTCAGAATGAGGGTAGTACAGATGAGAAGTC 923
     ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
   1 GluAsnGluGluGlnAspSerGlnAsnGluGlySerThrAspGluLysSe 17

 924 AAGCCCTGCCAGCTCTCAAGAAGGGTCTCCTTCAGGTGACCAGCAGTTTT 973
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  17 rSerProAlaSerSerGlnGluGlySerProSerGlyAspGlnGlnPheS 34

 974 CACCTAAATCCAACACTGAAAAATCTAAAGGGGAATTAATGTTTGATGAT 1023
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  34 erProLysSerAsnThrGluLysSerLysGlyGluLeuMetPheAspAsp 50

1024 TCTTCAGATTCATCACCGGAAAAACAGGAGAGAAATTTAAACTGGACCCC 1073
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  51 SerSerAspSerSerProGluLysGlnGluArgAsnLeuAsnTrpThrPr 67

1074 GGCCGAAGTCCCACAGTTAGCTGCAGCAAAACGCAGGCTGCCTCAGGGAA 1123
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  67 oAlaGluValProGlnLeuAlaAlaAlaLysArgArgLeuProGlnGlyL 84

1124 ACGAGCCTGGGTTGATTAACTTGTGTGCCAATGTCCCACCCGTCCCAGGT 1173
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  84 ysGluProGlyLeuIleAsnLeuCysAlaAsnValProProValProGly 100

1174 AACATTTTGCCCCCTGAGGTCCGGGGTAATTTAATGGCTGCTGGACAAAA 1223
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 101 AsnIleLeuProProGluValArgGlyAsnLeuMetAlaAlaGlyGlnAs 117

1224 CCTCCAAAGTTCTGAAAGATCAGAAATGATAGCTACCTGGAGTCCAGCTG 1273
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 117 nLeuGlnSerSerGluArgSerGluMetIleAlaThrTrpSerProAlaV 134

1274 TACGGACACTGAGGAATATTACTAATAATGCTGACATTCAGCAGATGAAC 1323
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 134 alArgThrLeuArgAsnIleThrAsnAsnAlaAspIleGlnGlnMetAsn 150

1324 CGGCCATCAAATGTAGCACATATCTTACAGACTCTTTCAGCACCTACGAA 1373
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 151 ArgProSerAsnValAlaHisIleLeuGlnThrLeuSerAlaProThrLy 167

1374 AAATTTAGAACAGCAGGTGAATCACAGCCAGCAGGGACATACAAATGCCA 1423
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 167 sAsnLeuGluGlnGlnValAsnHisSerGlnGlnGlyHisThrAsnAlaA 184

1424 ATGCAGTGCTGTTTAGCCAAGTGAAAGTGACTCCACAGACACACATGCTA 1473
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 184 snAlaValLeuPheSerGlnValLysValThrProGluThrHisMetLeu 200

1474 CAGCAGCAGCAGCAGGCCCAGCAGCAGCAGCAGCACCCGGTTTTACA 1523
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 201 GlnGlnGlnGlnGlnAlaGlnGlnGlnGlnGlnHisProValLeuHi 217

1524 CCTTCAGCCCCAGCAGATAATGCAGCTCCAGCAGCAGCAGCAGCAGACA 1573
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 217 sLeuGlnProGlnGlnIleMetGlnLeuGlnGlnGlnGlnGlnGlnGlnI 234

1574 TCTCTCAGCAACCTTACCCCCAGCAGCCGCCGCATCCATTTTCACAGCAG 1623
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 234 leSerGlnGlnProTyrProGlnGlnProProHisProPheSerGlnGln 250

1624 CAGCAGCAGCAACAGCAGCCACCACCATCGCCTCAGCAGCATCAGCTTTT 1673
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 251 GlnGlnGlnGlnGlnGlnProProProSerProGlnGlnHisGlnLeuPh 267

1674 TGGACATGATCCAGCAGTGGAGATTCCAGAAGAAGGCTTCTTATTGGGAT 1723
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 267 eGlyHisAspProAlaValGluIleProGluGluGlyPheLeuLeuGlyC 284

1724 GTGTGTTTGCAATTGCGGATTATCCAGAGCAGATGTCTGATAAGCAACTG 1773
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 284 ysValPheAlaIleAlaAspTyrProGluGlnMetSerAspLysGlnLeu 300

1774 CTGGCCACCTGGAAAAGGATAATCCAGGCACATGGCGGCACTGTTGACCC 1823
     ||||||||||||||||||||||||||||||||||||||||||||||||| |
 301 LeuAlaThrTrpLysArgIleIleGlnAlaHisGlyGlyThrValAsp.P 317

1824 CACCTTCACGAGTCGATGCACGCACCTTCTCTGTGAGAGTCAAGTCAGCA 1873
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 317 roProSerArgValAspAlaArgThrPheSerValArgValLysSerAla 333

1874 GCGCGTAT.GCACAGGCAATAAGAGAAAGAAAAGAGATGTGTTACTGCACA 1922
     ||||||   ||||||||||||||||||||||||||||||||||||||||||
 334 AlaArgIleAlaGlnAlaIleArgGluArgLysArgCysValThrAlaHi 350

1923 CTGGTTAAACACAGTCTTAAAAAAAAAAAAAAATGGTACCGCCGCACCGAG 1972
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 350 sTrpLeuAsnThrValLeuLysLysLysLysMetValProProHisArgA 367

1973 CCCTTCACTTCCCAGTGGCCTTCCCACCAGGAGGAAAGCCATGTTCACAG 2022
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 367 laLeuHisPheProValAlaPheProProGlyGlyLysProCysSerGln 383

2023 CATATTATTTCTGTGACTGGATTTGTTCATAGTGACAGAGATGACCTAAA 2072
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 384 HisIleIleSerValThrGlyPheValAspSerAspArgAspAspLeuLy 400

2073 ATTAATGGCTTATTTGGCAGGTGCCAAATATACGGGTTATCTATGCCGCA 2122
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 400 sLeuMetAlaTyrLeuAlaGlyAlaLysTyrThrGlyTyrLeuCysArgS 417

2123 GCAACACAGTCCTCATCTGTAAAGAACCAACTGGTTTAAAGTATGAAAAA 2172
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 417 erAsnThrValLeuIleCysLysGluProThrGlyLeuLysTyrGluLys 433

2173 GCCAAAGAGTGGAGGATACCCTGTGTCAACGCCCAGTGGCTTGGCGACAT 2222
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 434 AlaLysGluTrpArgIleProCysValAsnAlaGlnTrpLeuGlyAspIl 450

2223 TCTTCTGGGAAACTTTGAGGCACTGAGGCAGATTCAGTATAGTCGCTACA 2272
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 450 eLeuLeuGlyAsnPheGluAlaLeuArgGlnIleGlnTyrSerArgTyrT 467

2273 CGGCATTCAGTCTGCAGGATCCATTTGCCCCTACCCAGCATTTAGTTTTA 2322
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 467 hrAlaPheSerLeuGlnAspProPheAlaProThrGlnHisLeuValLeu 483

2323 AATCTTTTAGATGCTTGGAGAGTTCCCTTAAAAGTGTCTGCAGAGTTGTT 2372
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 484 AsnLeuLeuAspAlaTrpArgValProLeuLysValSerAlaGluLeuLe 500

2373 GATGAGTATAAGACTACCTCCCAAACTGAAACAGAATGAAGTAGCTAATG 2422
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 500 uMetSerIleArgLeuProProLysLeuLysGlnAsnGluValAlaAsnV 517

2423 TCCAGCCTTCTTCCCAAAGAGCCAGAATTGAAGACGTACCACCTCCCACT 2472
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 517 alGlnProSerSerGlnArgAlaArgIleGluAspValProProProThr 533

2473 AAAAAGCTAACTCCAGAATTGACCCCTTTTGTGCTTTTCACTGGATTCGA 2522
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 534 LysLysLeuThrProGluLeuThrProPheValLeuPheThrGlyPheGl 550

2523 GCCTGTCCAGGTTCAACAGTATATTAAGAAGCTCTACATTCTTGGTGGAG 2572
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 550 uProValGlnValGlnGlnTyrIleLysLysLeuTyrIleLeuGlyGlyG 567

2573 AGGTTGCGGAGTCTGCACAGAAGTGCACACACCTCATTGCCAGCAAAGTG 2622
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 567 luValAlaGluSerAlaGlnLysCysThrHisLeuIleAlaSerLysVal 583

2623 ACTCGCACCGTGAAGTTCCTGACGGCGATTTCTGTCGTGAAGCACATAGT 2672
     |||||||||:::||||||||:::|||||||||||||||||||||||||||
 584 ThrArgThrLeuLysPheLeuAlaAlaIleSerValValLysHisIleVa 600

2673 GACGCCAGAGTGGCTGGAAGAATGCTTCAGGTGTCAGAAGTTCATTGATG 2722
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 600 lThrProGluTrpLeuGluGluCysPheArgCysGlnLysPheIleAspG 617
```

```
2723 AGCAGAACTACATTCTCCGAGATGCTGAGGCAGAAGTACTTTTCTCTTTC 2772
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 617 luGlnAsnTyrIleLeuArgAspAlaGluAlaGluValLeuPheSerPhe 633

2773 AGCTTGGAAGAATCCTTAAAACGGGCACACGTTTCTCCACTCTTTAAGGC 2822
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 634 SerLeuGluGluSerLeuLysArgAlaHisValSerProLeuPheLysAl 650

2823 AAAATATTTTTACATCACACCTGGAATCTGCCCAAGTCTTTCCACTATGA 2872
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 650 aLysTyrPheTyrIleThrProGlyIleCysProSerLeuSerThrMetL 667

2873 AGGCAATCGTAGAGTGTGCAGGAGGAAAGGTGTTATCCAAGCAGC:ATCT 2922
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 667 ysAlaIleValGluCysAlaGlyGlyLysValLeuSerLysGlnProSer 683

2923 TTCCGGAAGCTCATGGAGCACAAGCAGAACTCGAGTTTGTCGGAAATAAT 2972
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 684 PheArgLysLeuMetGluHisLysGlnAsnSerSerLeuSerGluIleIl 700

2973 TTTAATATCCTGTGAAAATGACCTTCATTTATGCCGAGAATATTTTGCCA 3022
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 700 eLeuIleSerCysGluAsnAspLeuHisLeuCysArgGluTyrPheAlaA 717

3023 GAGGCATAGATGTTCACAATGCAGAGTTCGTTCTGACTGGAGTGCTCACT 3072
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 717 rgGlyIleAspValHisAsnAlaGluPheValLeuThrGlyValLeuThr 733

3073 CAAACGCTGGACTATGAATCATATAAGTTTAAC 3105
     |||||||||||||||||||||||||||||||||
 734 GlnThrLeuAspTyrGluSerTyrLysPheAsn 744

seq_name: sp_vertebrate:Q90WJ3

seq_documentation_block:
ID   Q90WJ3      PRELIMINARY;      PRT;  1256 AA.
AC   Q90WJ3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SWIFT.
GN   K14.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE=21260044; PubMed=11359898;
RA   Shimizu K., Bourillot P.Y., Nielsen S.J., Zorn A.M., Gurdon J.B.;
RT   "Swift Is a Novel BRCT Domain Coactivator of Smad2 in Transforming
RT   Growth Factor beta Signaling.";
RL   Mol. Cell. Biol. 21:3901-3912(2001).
DR   EMBL; AF172855; AAK55123.1; -.
SQ   SEQUENCE   1256 AA;  144187 MW;  E7795B12C1A42DE1 CRC64;


alignment_scores:
            Quality: 3425.50            Length:   1332
              Ratio:  4.112              Gaps:     20
 Percent Similarity:  62.538   Percent Identity: 53.829

alignment_block:
 US-09-664-641-10 x Q90WJ3   ..

 Align seg 1/1  to: Q90WJ3  from: 1  to: 1256

     165 CCTTCAAATCACCTCTAGGCACAGGTTATTCAGCTTCTCAAGGCTGGAAA 214
         |||   :::   |||      :::|||:::|||:::|||||||||||||||
      20 ProValGlyAspLeuAspProLysValValGlnMetLeuLysAlaGlyLy 36

 215 AGCGAAGGAAGTTTCCTACAATGCACTAGCCTCACACATAATCTCAGAGG 264
     |||||||||||||||||||||||||||||:::|||||||||:::||||
  36 sAlaLysGluValSerTyrAsnAlaLeuAlaThrHisIleIleAlaGluA 53

 265 ATGGGGACAATCCAGAGGTGGGAGAAGCTCGGGAAGTCTTTGACTTACCT 314
     |||||||||||||||||||||||||||||||||||||||||||||||||||
  53 spGlyAspAsnProGluValGlyGluAlaArgGluValPheAspLeuPro 69

 315 GTTGTAAAGCCTTCTTGGGTGATTCTGTCCGTTCAGTGTGGAACTCTTCT 364
     |||||||||||||||||||||||||||||||:::|||||||:::   ||
  70 ValValLysProSerTrpValIleLeuSerValArgCysGlyValArgLe 86

 365 GCCAGTAAATGGTTTTTCTCCAGAATCATGTCAGATTTTTTTTGGAATCA 414
     ||||   ||||||||||||||||||||   |||||||||||||||:::|
  86 uProGluAsnGlyPheSerProGluSerGlyGlnIlePhePheGlyValT 103

 415 CTGCCTGCCTTTCTCAGGGTGTTGATACAAGCTGGAGCTCTTTGTTGGAG 464
     |||||||||||||||| |||
 103 hrAlaCysLeuSerGln.Val..................................... 109

 465 TCTTCCAGAGCTCTCCCAAGGAGAGGTAGGGAAGGGAGCTTGTCCAGCAG 514

 109 ..... 109

 515 AAGTTGGGAAGCACAGAGATCATCTGCCTTCTTCTGACCCGGTATTGATG 564

 109 ..... 109

 565 CAGGCTGAGGCCTCTGTTGTAATGTGCTGGGTGTCATCTGAAGACAGAAG 614
                                      |||   :::|||||::
 110 ..............................SerProAspAspArgAs 115

 615 TGCCCTGTGGGCTTTGGTTACGTTCTATGGGGGAGATTGCCAGCTAACCC 664
     ::::|||||||||||||:::|||||||||||||||||||||||||||:::|
 115 nSerLeuTrpAlaLeuThrThrPheTyrGlyGlyAspCysGlnLeuSerL 132

 665 TCAATAAGAAATGCACGCATTTGATTGTTCCAGAGCCAAAGGGGGGGAAA 714
     |||||||||||||||||||||||||||||||||||||||||||||:::|||
 132 euAsnLysLysCysThrHisLeuIleValProGluProLysGlyAsnLys 148

 715 TACGAATGTGCTTTAAAGCGAGCAAGTATTAAAATTGTGACTCCTGACTG 764
     ||||||   |||:::::|||:::|||||||||||||||||||||||||||
 149 TyrGluTyrAlaPheGlnArgGlySerIleLysIleValThrProAspTr 165

 765 GGTTCTGGATTGCGTATCAGAGAAAACCAAAAAGGACGAAGCATTTTATC 814
     ||||||||||   |||||||||||||||||||||||||||:::||||
 165 pValLeuAspSerValSerGluLysThrLysLysAspGluAlaLeuTyrH 182

 815 ATCCTCGTCTGATTATTTATGAAGAGGAAGAAGAGGAAGAGGAAGAGGAG 864
     ||||||||||||||:::|||       |||||||||||||||||||||||
 182 isProArgLeuIleValTyr......GluGluGluGluGluGluGluGlu 196

 865 GAGGAAGTAGAAAATGAGGAACAAGATTCTCAGAATGAGGGTAGTACAGA 914
     |||||||:::|||||||||||||||:::|||||||||   ::::::   ||
 197 GluGluAlaGluAsnGluGluGlnGluSerGlnAsnGlySerAspGluAs 213

 915 TGAG.....AAGTCAAGCCCTGCCAGCTCTCAAGAAGGGTCTCCTT 955
     |:::      ||||||||||||:::::::|||:::::|||||||||
 213 pLysTyrSerSerLysSerSerProValThrSerArgAspGlySerProV 230

 956 CAGGTGACCAGCAGTTTTCACCTAAA...TCCAACACTGAAAAATCTAAA 1002
         :::::::::   |||||||||   |||:::::::::|||   |||
 230 alHisAsnArgGluSerSerProLysArgSerSerSerAspLysIleLys 246

1003 GGGGAATTAATGTTTGATGATTCTTCAGATTCATCACCGGAAAAACAGGA 1052
     :::|||||||||||||||||||||||||||||||||||||||||||||||||
 247 SerGluLeuMetPheAspAspSerSerAspSerSerProGluLysGlnGl 263

1053 GAGAAATTTAAACTGGACCCCGGCCGAAGTCCCACAGTTAGCTGCAGCAA 1102
     |||||||||||||||||||||||||||||||||||||:::::::::||||
 263 uArgAsnLeuAsnTrpThrProAlaGluValProGlnMetSerThrAlaL 280

1103 AACGCAGGCTGCCTCAGGGAAAGGAGCCTGGGTTGATTAACTTGTGTGCC 1152
```


280 ysArqGlnLeuHisGlnGly 286
1153 AATGTCCCACCCGTCCCAGGTAACATTTTGCCCCCTGAGGTCCGGGGTAA 1202
287 AlaProGly 289
1203 TTTAATGGCTGCTGGACAAAACCTCCAAAGTTCTGAAAGATCAGAAATGA 1252
290 SerGluArgProAspMetM 296
1253 TAGCTACCTGGAGTCCAGCTGTACGGACACTGAGGAATATTACTAATAAT 1302
296 etAlaPheArgSerProAlaValArgThrLeuArgAsnIleThrAsnSer 312
1303 GCTGACATTCAGCAGATGAACCGGCCATCAAATGTAGCACATATCTTACA 1352
313 AlaAspValGlnGlnValAsnArgProSerAsnValAlaHisIleLeuGl 329
1353 GACTCTTTCAGCACCTACGAAAAATTTAGAACAGCAGGTGAATCACAGCC 1402
329 nThrLeuSerAlaSerThrLysSerLeuGluGlnArgValAsnHisProG 346
1403 AGCAGGGACATACAAATGCCAATGCAGTGCTGTTTAGCCAAGTGAAA... 1449
346 lnGlnGlyHis.....ProAsnAlaValLeuPheGlyGlnValLysPro 360
1450 GTGACTCCACAGACA...CACATGCTACAGCAG.....CAGCA 1484
361 LeuThrSerGluAlaGlnHisLeuIleGlnGlnSerHisGlnProHisHi 377
1485 GCAGGCCCAGCAGCAGCAGCAGCAGCACCCGGTTTTACACCTTCAGCCCC 1534
377 sGlnGlnGlnGlnGlnGlnGlnAsnHisProIleValGlnLeuGlnProG 394
1535 AGCAGATAATGCAGCTCCAGCAGCAGCAGCAGCAG..... 1569
394 lnGlnLeuLeuGlnLeuGlnGlnGlnGlnGlnGlnIleThrGlnGlnVal 410
1570 .....CAGATCTCTCAGCAA..... 1584
411 PheProGlnHisGlnPheProGlnValAsnGlnGlnHisProPheThrGl 427
1585 ....CCTTACCCCCAGCAGCCGCCGCATCCA..... 1611
427 nLeuGlnPheProGlnGlnGlnLeuHisProGlnGlnGlnLeuHisArgP 444
1612 .....TTTTCACAGCAG..... 1623
444 roGlnGlnGlnThrIleGlnHisPheGlnGlnGlnHisAlaLeuGlnGln 460
1623 1623
461 GlnLeuHisGlnLeuGlnGlnGlnHisLeuGlnProLysProGlnThrLe 477
1623 1623
477 uGlnGlnAsnMetGlnGlnGlnAsnLeuGlnGlnProAsnLeuGlnGlnI 494
1623 1623
494 leLeuGlnHisGlnGlnThrIleGlnGlnThrProSerGlnGlnGlnAla 510
1623 1623
511 LeuGlnProAlaIleGlnGlnGlnGlnMetLeuGlnProAsnIleGlnGl 527
1623 1623
527 nGlnGlnThrLeuGlnSerAsnLeuGlnGlnGlnThrLeuGlnProIleI 544
1623 1623
544 leGlnGlnGlnGlnGlyLeuGlnGlnGlnAsnLeuGlnGlnSerLeuGln 560
1623 1623
561 GlnIleGlnGlnGlnMetGlnHisLeuThrProGlnGlnLysGlnGlnIl 577
1624 .CAGCAGCAGCAACAGCAG..... 1641
577 eGlnGlnGlnGlnGlnGlnMetLeuGlnLysGlnGlnLeuGlnSerGlnS 594
1641 1641
594 erLeuGlnGlnGlnGlnMetGlnThrHisValLeuGlnGlnGlnGlnIle 610
1641 1641
611 GlnThrGlnAlaLeuGlnGlnGlnGlnGlnIleGlnAsnGlnAlaLeuGl 627
1641 1641
627 nGlnProGlnGlnAsnGlnValGlnThrHisAlaLeuGlnGlnHisArgL 644
1641 1641
644 euGlnSerGlnThrLeuGlnGlnGlnHisHisValGlnGlyGlnThrVal 660
1641 1641
661 GlnGlnGlnThrHisGlnLeuGlnThrGlnThrLeuGlnGlnGlnHisGl 677
1642 .....CCACCACCATCG..... 1653
677 nIleGlnThrLeuGlnIleProHisGlnIleProAlaProAsnGlnGlnH 694
1653 1653
694 isGlnIleProProGlnMetLeuGlnGlnGlnThrLeuGlnLeuGlnGln 710
1653 1653
711 GlnMetGlnProGlnIleGlnGlnProGlnMetGlnSerGlyValGlnGl 727
1653 1653
727 nGlnSerSerLeuGlnProGlnGlnMetGlnGlnHisLysHisAsnLeuG 744
1653 1653
744 lnGlnValGlnHisGlnLeuGlnGlnLeuGlnGlnGlnArgMetGlnGln 760
1654 .....CCTCAGCA 1661
761 ArgGlnMetAlaLeuProGlnGlnIleAlaAsnGlnGlnProProGlnPr 777
1662 GCATCAGCTTTTTGGACATGATCCAGCAGTGGAGATTCCAGAAGAAGGCT 1711
777 oHisGlnLeuHisGlyHisAspProSerValGluIleProGluAspPheP 794
1712 TCTTATTGGGATGTGTGTTTGCAATTGCGGATTATCCAGAGCAGATGTCT 1761
794 heLeuLeuGlyCysValPheAlaIleAlaAspTyrProGluGlnMetPro 810
1762 GATAAGCAACTGCTGGCCACCTGGAAAAGGATAATCCAGGCACATGGCGG 1811
811 AspLysGlnLeuLeuAlaThrTrpLysArgIleIleGlnThrHisGlyGl 827
1812 CACTGTTGACCCCACCTTCACGAGTCGATGCACGCACCTTCTCTGTGAGA 1861
827 yThrValAspAlaAlaLeuSerSerArgCysThrHisLeuLeuCysGluS 844
1862 GTCAAGTCAGCAGCGCGTATGCACAGGCAATAAGAGAAAGAAAGAGATGT 1911
844 erGlnValSerSerMetTyrAlaGlnAlaLeuLysGluArgLysArgCys 860

```
1912 GTTACTGCACACTGGTTAAACACAGTCTTAAAAAAAAAAAAAAATGGTACC 1961
     :::||||||||||||||||||:::||||||||||||||||||||||||||||
 861 IleThrAlaHisTrpLeuAsnAlaValLeuLysLysLysLysMetValPr 877

1962 GCCGCACCGAGCCCTTCACTTCCCAGTGGCCTTCCCACCAGGAGGAAAGC 2011
     ||||:::|||||||||||||||||||||||||||||||||||||||||||
 877 oProTyrArgAlaLeuHisPheProValAlaPheProProGlyGlyLysP 894

2012 CATGTTCACAGCATATTATTTCTGTGACTGGATTTGTTGATAGTGACAGA 2061
     ||||||||||||||||||||||||||||||||||||||||||.||||||
 894 roCysSerGlnHisIleIleSerValThrGlyPheValAspSerAspArg 910

2062 GATGACCTAAAATTAATGGCTTATTTGGCAGGTGCCAAATATACGGGTTA 2111
     ||||||||||||||||||:::|||||||||||||||||||||||||||||
 911 AspAspLeuLysLeuMetGlyTyrLeuAlaGlyAlaLysTyrThrGlyTy 927

2112 TCTATGCCCGCAGCAACACAGTCCTCATCTGTAAAGAACCAACTGGTTTAA 2161
     ||||||||||||||||||||||||||||||||||||||:::|||||||||
 927 rLeuCysArgSerAsnThrValLeuIleCysLysGluProSerGlyLeuL 944

2162 AGTATGAAAAAGCCAAAGAGTGGAGGATACCCTGTGTCAACGCCCAGTGG 2211
     ||||||||||||||||||||||||||||||||||||||||   |||
 944 ysTyrGluLysAlaLysGluTrpArgIleProCysValAsnAlaLeuTrp 960

2212 CTTGGCGACATTCTTCTGGGAAACTTTGAGGCACTGAGGCAGATTCAGTA 2261
     |||   ||||||||||||||||||:::|||||||||||||||||||||::
 961 LeuCysAspIleLeuLeuGlyAsnLeuGluAlaLeuArgGlnIleGlnHi 977

2262 TAGTCGCTACACGGCATTCAGTCTGCAGGATCCATTTGCCCCTACCCAGC 2311
     :|||||||||||||:::|||:::||||||||||||:::|||||||:::  |
 977 sSerArgTyrThrValPheAsnLeuGlnAspProLeuAlaProSerProH 994

2312 ATTTAGTTTTA...AATCTTTTAGATGCTTGGAGAGTTCCCTTAAAAGTG 2358
     ||||||||      :::|||||||||||||||||||:::|||||||||||
 994 isLeuValThrAspAspLeuLeuAspAlaTrpArgMetProLeuLysVal 1010

2359 TCTGCAGAGTTGTTGATGAGTATAAGACTACCTCCCAAACTGAAACAGAA 2408
     |||:::|||:::|||||||||||||||:::|||   |||   ||||||||
1011 SerSerGluValLeuMetSerIleArgMetProLeuLysProLysGlnAs 1027

2409 TGAAGTAGCTAATGTCCAGCCTTCTTCCCAAAGAGCCAGAATTGAAGACG 2458
     ||||   |||   |||||||||      :::|||   ||||||||||||:
1027 nGluProAla...ValGlnPro......LysArgProArgIleGluAspI 1041

2459 TACCACCTCCCACTAAAAAGCTAACTCCAGAATTGACCCCTTTTGTGCTT 2508
     ::|||||||||||||||||||||:::|||:::   ||||||   |||:::
1041 leProProProThrLysLysLeuSerProAspGlnThrProHisValIle 1057

2509 TTCACTGGATTCGAGCCTGTCCAGGTTCAACAGTATATTAAGAAGCTCTA 2558
     ||||||||||||:::|||:::|||||||||||||||||||||||||||||
1058 PheThrGlyPheAspProLeuGlnValGlnGlnTyrIleLysLysLeuTy 1074

2559 CATTCTTGGTGGAGAGGTTCCGGAGTCTGCACACAAGTGCACACACCTCA 2608
     ||||||||||||||||||||||:::::|||||||||||||||||||||:
1074 rIleLeuGlyGlyGluValAlaAspThrAlaGlnLysCysThrHisLeuV 1091

2609 TTGCCAGCAAAGTGACTCGCACCGTGAAGTTCCTGACGGCGATTTCTGTC 2658
     ::||||:::|||||||||||||||||||||||||||||||||||||||||
1091 alAlaAsnLysValThrArgThrValLysPheLeuThrAlaIleSerVal 1107

2659 GTGAAGCACATAGTGACGCCAGAGTGGCTGGAAGAATGCTTCAGGTGTCA 2708
     :::|||||||||||||||||||||||||||:::|||   |||:::   ||
1108 AlaLysHisIleValThrProGluTrpLeuAspGluSerPheLysSerGl 1124

2709 GAAGTTCATTGATGAGCAGAACTACATTCTCCGAGATGCTGAGGCAGAAG 2758
     |||||||   :::|||||||||||||||||||||||||||||||||||||
1124 nLysPheAlaGluGluGlnAsnTyrIleLeuArgAspAlaGluAlaGluV 1141

2759 TACTTTTCTCTTTCAGCTTGGAAGAATCCTTAAAACGGGCACACGTTTCT 2808
     ||||||||   ||||||||||||||||||||||||:::||||||||:::
1141 alLeuPheCysPheSerLeuGluGluSerLeuLysLysAlaHisValAsn 1157
```

```
2809 CCACTCTTTAAGGCAAAATATTTTTACATCACACCTGGAATCTGCCCAAG 2858
     ||||||||||||:::|||||||||||||||||||||||||||||||||||
1158 ProLeuPheLysGlyLysTyrPheTyrIleThrProGlyIleCysProSe 1174

2859 TCTTTCCACTATGAAGGCAATCGTAGAGTGTGCAGGAGGAAAGGTGTTAT 2908
     ||||||||||||||||||||||||||||||||||||||||||||:::|||:
1174 rLeuSerThrMetLysAlaIleValGluCysAlaGlyGlyLysIleLeuT 1191

2909 CCAAGCAGCCATCTTTCCGGAAGCTCATGGAGCACAAGCAGAACTCGAGT 2958
     ::|||||||||||||||||||:::|||||||||||||||||||:::
1191 hrLysGlnProSerPheArgLysIleMetGluHisLysGlnAsnLysArg 1207

2959 TTGTCGGAAATAATTTTAATATCCTGTGAAAATGACCTTCATTTATGCCG 3008
     |||:::||||||||||||||||||||||||||||||||||||||||||||
1208 LeuAlaGluIleIleLeuIleSerCysGluAsnAspLeuHisLeuCysAr 1224

3009 AGAATATTTTGCCAGAGGCATAGATGTTCACAATGCAGAGTTCGTTCTGA 3058
     ||||||||||||   ::::::||||||||||||||||||||||||||||
1224 gGluTyrPheAlaGlySerValAspValHisAsnAlaGluPheValLeuT 1241

3059 CTGGAGTGCTCACTCAAACGCTGGACTATGAATCATATAAGTTT 3102
     ||||||||||||||||:::||||||||||||||||||||||||
1241 hrGlyValLeuThrGlnAlaLeuAspTyrGluSerTyrLysPhe 1255
```

seq_name: sp_human:Q96HP2

seq_documentation_block:

```
ID   Q96HP2      PRELIMINARY;      PRT;   391 AA.
AC   Q96HP2;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   UNKNOWN (PROT IN FOR IMAGE:3503689) (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN, AND NEUROBLASTOMA;
RA   Strausberg R.;
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC008328; AAH08328.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   391 AA;  44356 MW;  4E4D7562F613C834 CRC64;
```

alignment_scores:

```
            Quality: 2039.00              Length:    391
              Ratio:  5.215                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 99.744
```

alignment_block:

US-09-664-641-10 x Q96HP2 ..

Align seg 1/1 to: Q96HP2 from: 1 to: 391

```
   1933 ACAGTC TAAAAAAAAAAAAAAATGGTACCGCCGCACCGAGCCCTTCACTT 1982
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 ThrValLeuLysLysLysLysMetValProProHisArgAlaLeuHisPh 17

   1983 CCCAGTGGCCTTCCCACCAGGAGGAAAGCCATGTTCACAGCATATTATTT 2032
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     17 eProValAlaPheProProGlyGlyLysProCysSerGlnHisIleIleS 34

   2033 CTGTGACTGGATTTGTTGATAGTGACAGAGATGACCTAAAATTAATGGCT 2082
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     34 erValThrGlyPheValAspSerAspArgAspAspLeuLysLeuMetAla 50

   2083 TATTTGGCAGGTGCCAAATATACGGGTTATCTATGCCGCAGCAACACAGT 2132
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     51 TyrLeuAlaGlyAlaLysTyrThrGlyTyrLeuCysArgSerAsnThrVa 67
```

```
2133 CCTCATCTGTAAAGAACCAACTGGTTTAAAGTATGAAAAAGCCAAAGAGT 2182
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  67 lLeuIleCysLysGluProThrGlyLeuLysTyrGluLysAlaLysGluT  84

2183 GGAGGATACCCTGTGTCAACGCCCAGTGGCTTGGCGACATTCTTCTGGGA 2232
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  84 rpArgIleProCysValAsnAlaGlnTrpLeuGlyAspIleLeuLeuGly 100

2233 AACTTTGAGGCACTGAGGCAGATTCAGTATAGTCGCTACACGGCATTCAG 2282
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 101 AsnPheGluAlaLeuArgGlnIleGlnTyrSerArgTyrThrAlaPheSe 117

2283 TCTGCAGGATCCATTTGCCCCTACCCAGCATTTAGTTTTAAATCTTTTAG 2332
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 117 rLeuGlnAspProPheAlaProThrGlnHisLeuValLeuAsnLeuLeuA 134

2333 ATGCTTGGAGAGTTCCCTTAAAAGTGTCTGCAGAGTTGTTGATGAGTATA 2382
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 134 spAlaTrpArgValProLeuLysValSerAlaGluLeuLeuMetSerIle 150

2383 AGACTACCTCCCAAACTGAAACAGAATGAAGTAGCTAATGTCCAGCCTTC 2432
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 151 ArgLeuProProLysLeuLysGlnAsnGluValAlaAsnValGlnProSe 167

2433 TTCCCAAAGAGCCAGAATTGAAGACGTACCACCTCCCACTAAAAAGCTAA 2482
     ||||:::||||||||||||||||||||||||||||||||||||||||||
 167 rSerLysArgAlaArgIleGluAspValProProProThrLysLysLeuT 184

2483 CTCCAGAATTGACCCCTTTTGTGCTTTTCACTGGATTCGAGCCTGTCCAG 2532
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 184 hrProGluLeuThrProPheValLeuPheThrGlyPheGluProValGln 200

2533 GTTCAACAGTATATTAAGAAGCTCTACATTCTTGGTGGAGAGGTTGCGGA 2582
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 201 ValGlnGlnTyrIleLysLysLeuTyrIleLeuGlyGlyGluValAlaGl 217

2583 GTCTGCACAGAAGTGCACACACCTCATTGCCAGCAAAGTGACTCGCACCG 2632
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 217 uSerAlaGlnLysCysThrHisLeuIleAlaSerLysValThrArgThrV 234

2633 TGAAGTTCCTGACGGCGATTTCTGTCGTGAAGCACATAGTGACGCCAGAG 2682
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 234 alLysPheLeuThrAlaIleSerValValLysHisIleValThrProGlu 250

2683 TGGCTGGAAGAATGCTTCAGGTGTCAGAAGTTCATTGATGAGCAGAACTA 2732
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 251 TrpLeuGluGluCysPheArgCysGlnLysPheIleAspGluGlnAsnTy 267

2733 CATTCTCCGAGATGCTGAGGCAGAAGTACTTTTCTCTTTCAGCTTGGAAG 2782
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 267 rIleLeuArgAspAlaGluAlaGluValLeuPheSerPheSerLeuGluG 284

2783 AATCCTTAAAACGGGCACACGTTTCTCCACTCTTTAAGGCAAAATATTTT 2832
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 284 luSerLeuLysArgAlaHisValSerProLeuPheLysAlaLysTyrPhe 300

2833 TACATCACACCTGGAATCTGCCCAAGTCTTTCCACTATGAAGGCAATCGT 2882
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 301 TyrIleThrProGlyIleCysProSerLeuSerThrMetLysAlaIleVa 317

2883 AGAGTGTGCAGGAGGAAAGGTGTTATCCAAGCAGCCATCTTTCCGGAAGC 2932
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 317 lGluCysAlaGlyGlyLysValLeuSerLysGlnProSerPheArgLysL 334

2933 TCATGGAGCACAAGCAGAACTCGAGTTTCTCGGAAATAATTTTAATATCC 2982
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 334 euMetGluHisLysGlnAsnSerSerLeuSerGluIleIleLeuIleSer 350

2983 TGTGAAAATGACCTTCATTTATGCCGAGAATATTTTGCCAGAGGCATAGA 3032
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 351 CysGluAsnAspLeuHisLeuCysArgGluTyrPheAlaArgGlyIleAs 367

3033 TGTTCACAATGCAGAGTTCGTTCTGACTGGAGTGCTCACTCAAACGCTGG 3082
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 367 pValHisAsnAlaGluPheValLeuThrGlyValLeuThrGlnThrLeuA 384

3083 ACTATGAATCATATAAGTTTAAC 3105
     ||||||||||||||||||||||
 384 spTyrGluSerTyrLysPheAsn 391
```

seq_name: sp_invertebrate:Q9VUB6

seq_documentation_block:

```
ID   Q9VUB6      PRELIMINARY;      PRT;  1798 AA.
AC   Q9VUB6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CG8797 PROTEIN.
GN   CG8797.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003536; AAF49772.1; -.
DR   FlyBase; FBgn0036399; CG8797.
DR   InterPro; IPR001357; BRCT.
DR   Pfam; PF00533; BRCT; 4.
DR   SMART; SM00292; BRCT; 4.
DR   PROSITE; PS50172; BRCT; 2.
SQ   SEQUENCE   1798 AA;  205695 MW;  64E4963B1181B6E9 CRC64;
```

```
alignment_scores:
           Quality:  852.00            Length:    848
             Ratio:  1.906               Gaps:     24
 Percent Similarity:  52.712   Percent Identity: 28.774

alignment_block:
 US-09-664-641-10 x Q9VUB6   ..

 Align seg 1/1  to: Q9VUB6  from: 1  to: 1798

     925 AGCCCTGCCAGCTCTCAAGAAGGGTCTCCTTCAGGTGACCAGCAGTTTTC 974
         ||||||   :::::::|||   |||:::         :::|||||||   ||
     974 SerProIleGlnGlnGlnProGlyAlaGlyProGlnGlnGlnGlnGlnSe 990

     975 ACCTAAATCCAACACTGAAAAATCTAAAGGGGAATTAATGTTTGATGATT 1024
         ||||:::::::::      :::   :::                |||   :
     990 rProGlnGlnGlnProPheAsnProAsn.............AspProA 1002

    1025 CTTCAGATTCATCACCGGAAAAACAGGAGAGAAATTTAAACTGGACCCCG 1074
         ::::::::          ::::::|||:::   :::::::::      |||
    1002 snGlnGlnMetLeuMetGlnArgGlnGlnLeuArgValGlnGlnIlePro 1018

    1075 GCCGAA.........GTCCCACAGTTAGCTGCAGCAAAACGCAGGCTGCC 1115
            :::               ||||||:::           :::|||
    1019 IleGlnGlnGlnGlyProGlnGlnLeuGlyLysGlnThrValGlnLeuIl 1035

    1116 TCAGGGAAACGAGCCTGGGTTGATT..................... 1140
             |||   :::      ||||||
    1035 eThrGlyProAsnGlyGlnLeuIleGluGlnGlnThrIleValGlnGlnP 1052

    1141 ........AACTTGTGTGCCAATGTCCCACCCGTCCCA......GGTAAC 1176
                 :::   :::   :::::::|||   :::|||      |||:::
    1052 roGlnGlnGlnSerAlaProGlyThrProThrThrProGlyAlaGlyGly 1068

    1177 ATTTTGCCCCCTCAGGTCCGGGGTAATTTAATGGCTGCTGGACAAAACCT 1226
         :::|||||           ||||||              |||:::||
    1069 ValLeuPro.............GlyAsnAlaAsnIleMetThrGlnThrLe 1081

    1227 C...CAAAGTTCTGAAAGATCAGAAATGATAGCTACCTGGAGTCCAGCTG 1273
         |   ::::::|||         :::               ||||||
    1081 uValMetThrSerThrThrProAspGlyGlnProGlnGlnSerProGlnG 1098

    1274 TACGGACACTGAGGAATATTACTAATAATGCTGACATTCAGCAGATGAAC 1323
            :::|||::::::   |||      |||::::::   :::   :::
    1098 lnMetAsnLeuLysThrLysThrAlaLeuAlaAsnMetLeuSerAsnArg 1114

    1324 CGGCCATCAAATGTAGCACATATCTTACAGACTCTTTCAGCACCTACGAA 1373
            :::|||   |||:::      |||
    1115 LeuGlyAsnAsnGlyAlaGlnThrProGln..................... 1124

    1374 AAATTTAGAACAGCAGGTGAATCACAGCCAGCAGGGACATACAAATGCCA 1423
                 ||||||:::   :::      :::|||:::         :
    1125 .........GlnGlnIleIleGlnLeuProAspGlyGlnGlnProGlnG 1138

    1424 ATGCAGTGCTGTTTAGCCAAGTGAAAGTGACT.............CCA 1458
         ::      :::   :::||||||   |||:||
    1138 lnGlnGlnIleValGlnGlnValValValThrGlyAlaIleGlnGlnGln 1154

    1459 GAGACACACATGCTACAGCAGCAGCAGCAGGCCCAGCAGCAGCAGCAGCA 1508
         :::   ::::::|||   ||||||||||||   ||||||||||||||||
    1155 GlnGlnGlnLeuLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGl 1171

    1509 G.....................CACCCGGTTTTACACCTTCAGCCC.... 1533
         |                      |||      |||::::::|||
    1171 nMetValValGlnGlnGlnLeuProProGlnGlnHisValGluProProS 1188

    1533 ..................................................... 1533

    1188 erAlaAlaGlyAlaLeuArgMetMetGlyGlnGlnHisAsnAlaThrAla 1204

    1534 ........................CAGCAGATAATGCAGCTCCAGCA 1556
                                 |||:::::::::::::|||||||
    1205 GlyValProGlyProProArgThrGlnGlnGluLeuLeuMetLeuGlnGl 1221

    1557 GCAGCAGCAGCAGCAGATCTCTCAGCAA..................... 1584
         ||||||||||||||||:::   ||||||
    1221 nGlnGlnGlnGlnGlnLeuLeuGlnGlnGlnGlnMetArgArgValValA 1238

    1585 .....CCTTACCCCCAGCAGCCGCCGCATCCATTT.............. 1614
              |||      ||||||   |||:::
    1238 laLeuProGlnGlnGlnGlnAlaProGlnGlnHisLeuValGlnGlnGln 1254

    1615 .....................TCACAGCAGCAGCAGCAGCAA........ 1635
                               |||   ||||||||||||||||
    1255 IleValValAlaProThrGlnSerProGlnGlnGlnGlnGlnMetAlaVa 1271

    1635 ..................................................... 1635

    1271 lGlyValGlyValGlyValProValGlnArgThrProHisGlyTyrIleG 1288

    1636 .............CAGCAGCCACCACCATCGCCTCAGCAGCATCAGCTT 1671
                     |||||||||   |||   |||         |||:::
    1288 lnThrArgProGlyGlnGlnProIleProMetPro.........GlnPhe 1301

    1672 TTTGGACATGATCCAGCAGTGGAGATTCCAGAAGAAGGCTTCTTATTGGG 1721
         :::|||||:::|||   :::::::::|||   :::   ||||||:::||
    1302 TyrGlyHisAsnProAsnLeuLysLeuProAlaAspLeuPheLeuValGl 1318

    1722 ATGTGTGTTTGCAATTGCGGATTATCCAGAGCAGATGTCTGATAAGCAAC 1771
         ||||:::|||   |||:::::::|||   |||      :::|||:::
    1318 yCysThrPheTyrIleValGluTyrAspGluThrAspGlyAspGlu.... 1333

    1772 TGCTGGCCACCTGGAAAAGGATAATCCAGGCACATGGCGGCACTGTTGAC 1821
           |||      |||         |||:::      ||||||   ::::::
    1334 ..LeuProIleTrpLeuAspThrIleArgGlnPheGlyGlyAspIleGlu 1349

    1822 CCCACCTTCACGAGTCGATGCACGCACCTTCTCTGTGAGAGTCAAGTCAG 1871
         ::::::      |||   ||||||::::::|||::::::|||
    1350 ArgValTyrCysProArgValThrHisValIleCysArgThrGlnArgHi 1366

    1872 CAGCGCGTATGCACAGGCAATAAGAGAAAGAAAGAGATGTGTTACTGCAC 1921
          ::::::      ||||||:::|||:::   ||||||||||||||||||:
    1366 sGlyValValMetGlnAlaLeuArgAspAlaLysArgCysValThrAlaT 1383

    1922 ACTGGTTAAACACAGTCTTAAAAAAAAAAAAAAATGGTACCGCCGCACCGA 1971
         ::|||||:::   :::      |||:::::::::::::||||||   :::
    1383 yrTrpLeuSerAspIleCysLeuLysArgGlnLeuMetProProTrpGln 1399

    1972 GCCCTTCACTTCCCAGTGGCCTTCCCACCAGGA.........GGAAAGCC 2012
         |||||||:::|||      ||||||                  |||||
    1400 ProLeuHisLeuPro......PheProSerGlnPheGlyTyrArgLysPr 1414

    2013 ATGTTCACAGCATATTATTTCTGTGACTGGATTTGTTGATAGTGACAGAG 2062
         |   :::::::::||||||:::      ||||||      ::::::
    1414 oLeuGluArgTyrIleIleThrSerGluGlyPheGluGlyGluGluValV 1431

    2063 ATGACCTAAAATTAATGGCTTATTTGGCAGGTGCCAAATATACGGGTTAT 2112
              |||:::   ||||||      :::|||||||   ||||||:::|||
    1431 alArgLeuGlnGlnMetAlaGluGluCysGlyAlaIleTyrThrSerTyr 1447

    2113 CTATGCCGCAGCAACACAGTCCTCATCTGTAAAGAACCAACTGGTTTAAA 2162
         |||   :::   |||||||||:::::|||||:::      |||   ||
    1448 LeuSerLysValAsnThrValValValCysLysGlnLeuGluGlyAsnLy 1464

    2163 GTATGAAAAAGCCAAAGAGTGGAGGATACCCTGTGTCAACGCCCAGTGGC 2212
         |::::::   ||||||||||||:::||||||   |||||||||   ||||
    1464 sPheAsnAlaAlaLysGluTrpAsnIleProMetValAsnAlaLeuTrpL 1481

    2213 TTGGCGACATTCTTCTGGGAAACTTTGAGGCACTGAGGCAGATTCAGTAT 2262
         ||:::|||:::   :::||||||:::::::::|||   |||   :::
    1481 euSerAspValCysIleGlyAsnLeuSerGlyLeuSerGlnTyrGluAsn 1497

    2263 AGTCGCTACACGGCATTCAGTCTGCAGGATCCATTTGCCCCTACCCAGCA 2312
         :::|||      ::::::|||      ||||||           ::
    1498 ProLysTyrGlnGlnTyrAsnLeuValAlaProPheArgIleGluCysAs 1514
```

```
 2313 TTTAGTTTTAAATCTTTTAGATGCTTGGAGAGTTCCCTTAAAAGTGTCTG 2362
      :||||||   :::|||||||   ||||||:::::::|||:::::::::::::
 1514 nLeuValAlaHisLeuLeuThrAlaTrpLysAlaProIleAsnLeuThrG 1531

 2363 CAGAGTTGTTGATGAGTATAAGACTA..................... 2388
        |||           ::::::
 1531 lnGluAlaHisGluArgValLysArgHisLeuSerAspProTyrGlyAsn 1547

 2389 CCTCCCAAACTGAAACAGAATGAAGTAGCTAATGTCCAGCCTTCTTCCCA 2438
           |||||||||:::::::::::          |||   ::::::
 1548 GluGlnLysLeuLysArgGlnLysMetAsnAlaPheGlnGlnGluGlnLe 1564

 2439 AAGAGCCAGAATTGAAGACGTACCACCTCCCACTAAAAAGCTAACTCCAG 2488
            :::|||      |||      ||||||            |||:
 1564 uProGluGlnIleValCysValGluTyrProThrThrThrLysProProL 1581

 2489 AATTGACCCCTTTTGTGCTTTTCACTGGATTCGAGCCTGTCCAGGTTCAA 2538
      :::::                      ||||::  |||
 1581 ysValIle.....................PheSerGlnValAlaAspAla 1590

 2539 CAGTATATTAAGAAG...CTCTACATTCTTGGTGGAGAGGTTGCGGAGTC 2585
      :::   :::||||||   :::   ||||||||||||   |||::::::||
 1591 GluAlaLeuLysLysAlaValLeuIleLeuGlyGlyIleValValAspSe 1607

 2586 TGCACAGAAGTGCACACACCTCATTGCCAGCAAAGTGACTCGCACCGTGA 2635
      |         :::||||||||||:::   ::::::   :::||||||   |
 1607 rProAlaAspAlaThrHisLeuValMetThrArgGluSerArgThrCysL 1624

 2636 AGTTCCTGACGGCGATTTCTGTCGTGAAGCACATAGTGACGCCAGAGTGG 2685
      ||::::::   |||         |||   :::::::::      :::|||
 1624 ysLeuIleGlnAlaCysCysHisValAspTyrValLeuLysSerSerTrp 1640

 2686 CTGGAAGAATGCTTCAGGTGTCAGAAGTTCATTGATGAGCAGAACTACAT 2735
      :::   :::      ::::::   |||||:::      :::   |||
 1641 IleAlaAspSerAlaLysAlaGlyLysPheValProThrAspProTyrAr 1657

 2736 TCTCCGAGATGCTGAGGCAGAAGTACTTTTCTCTTTCAGCTTGGAAGAAT 2785
       ::::::         ::::::      :::::::|||:::|||:::
 1657 gIleGlnHisIleProValAspGluAsnLeuGlnPheAsnLeuAsnThrV 1674

 2786 CCTTAAAACGGGCACACGTTTCTCCACTCTTTAAGGCAAAATATTTTTAC 2835
        |||                |||   ||||||   :::||||||||:::
 1674 alLeuCysAlaProThrArgSerThrLeuPheAlaGlyLysTyrPheHis 1690

 2836 ATCACACCTGGAATCTGCCCAAGTCTTTCCACTATGAAGGCAATCGTAGA 2885
      :::||||||   :::   |||:::   :::   :::      ::::::||
 1691 ValThrProAspValPheProAlaArgGluGluIleIleArgMetIleGl 1707

 2886 GTGTGCAGGAGGAAAGGTGTTATCCAAGCAGCCATCTTTCCGGAAGCTCA 2935
      |   :::||||||||||||      |||:::   |||      ::::::
 1707 uSerSerGlyGlyLysValGluProLysArgArgSerGlyAlaSerValA 1724

 2936 TGGAG...CACAAGCAGAACTCGAGTTTGTCGGAAATAATTTTAATATCC 2982
        |||   |||   |||      :::   ::::::   |||:::::::::
 1724 laGluThrHisMetGln......AlaProAspSerTyrIleIleValThr 1738

 2983 TGTGAAAATGACCTTCATTTATGCCGAGAATATTTTGCCAGAGGC..... 3027
      |||   :::|||:::||||||||   :::        :::|||
 1739 CysProThrAspMetHisLeuCysAlaAspLeuThrArgHisGlyAsnPr 1755

 3028 ....ATAGATGTTCACAATGCAGAGTTCGTTCTGACTGGAGTGCTCACTC 3073
             :::   ::::::||||||||:::::::::::|||   |
 1755 oLysCysHisIleValSerThrGluPheValMetSerSerIleLeuArgG 1772

 3074 AAACGCTGGACTATGAA............TCATATAAGTTTAAC 3105
      ||   |||:::   |||            |||   :::|||
 1772 lnGlnLeuGluIleGluProAsnLeuIleProTyrLeuTyrAsn 1786
```

seq_name: sp_invertebrate:Q94046

seq_documentation_block:

```
ID   Q94046      PRELIMINARY;      PRT;  1074 AA.
AC   Q94046;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   T13F2.3A PROTEIN.
GN   T13F2.3A.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Swinburne J.;
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; Z81122; CAB03354.1; -.
DR   InterPro; IPR001357; BRCT.
DR   Pfam; PF00533; BRCT; 4.
DR   SMART; SM00292; BRCT; 4.
DR   PROSITE; PS50172; BRCT; 4.
SQ   SEQUENCE   1074 AA;  120339 MW;  F52E142322610C50 CRC64;
```

alignment_scores:

```
           Quality:  524.00            Length:    692
             Ratio:  1.444               Gaps:     24
 Percent Similarity:  52.457   Percent Identity: 25.000
```

alignment_block:

US-09-664-641-10 x Q94046 ..

Align seq 1/1 to: Q94046 from: 1 to: 1074

```
 1350 ACAGACTCTTTCAGCACC......TACGAAAAATTTAGAACAGCAGGTGA 1393
      |||   :::|||||||:::      ::::::   ||||||   :::
  347 ThrProThrPheSerSerProAlaHisGlnMetPheArgGlnGlyGlnPr 363

 1394 ATCACAGCCAGCAGGGACATACAAATGCCAATGCAGTGCTGTTTAGCCAA 1443
       :::   |||   |||      :::              :::   |||:
  363 oThrAlaProProGlyMetArgGln.....TyrIleSerProG 376

 1444 GTGAAAGTGACTCCAGAGACACACATGCTACAGCA.........GCAGCA 1484
      ::   ::::::::::::  . |||   :::|||            |||
  376 lnAlaGlyGlnGlnGlnProThrProGlyThrProHisArgPheAlaPro 392

 1485 GCAGGCCCAGCAGCAGCAGCAGCAGCACCCGGTTTTACACCTTCAGCCCC 1534
        ||||||:::::::::::::::::::|||   |||:::|||:::|||||
  393 ProGlyProSerGlyThrThrSerSerProHisPheAlaProGlyAlaPr 409

 1535 AGCAGATAAT.................GCAGCTCCAGCAGCAGCAGCAG 1566
      |:::   :::                 :::||| :::||||||:::
  409 oGlyAlaAspLeuArgSerProSerLeuMetSerPro.AsnGlnGlnSer 425

 1567 CAGCAG..................................................... 1572
      ||||||
  426 GlnGlnHisThrProValLeuProArgSerLeuHisAsnGluMetThrSe 442

 1573 .....................ATCTCTCAGCAACCTTACCCCCAGCAG. 1599
                          :::   :::||||||   ||||||
  442 rHisGlyMetArgProIleAlaLeuProSerGlnProLeuProGlnThrG 459

 1600 .....................................CCGCCGCAT... 1608
                                          |||||||:::
  459 lnGlyThrProGlnProIleLeuSerAlaSerSerLeuProProGlnSer 475

 1609 ............CCATTTTCACAGCAGCAGCAGCAGCAACAGCAGCCACC 1646
                     |||      :::   ||||||   :::|||
  476 AlaGlyAlaThrGlnPheValTyrArgProGlnGlnProAsnGlnMetAs 492
```

```
1647 ACCATCGCCTCACCAGCATCAGCTTTTTGGACATGATCCA..... 1686
             |||:::::::|||:::      ||||||
 492 nThrMetAsnGlnMetAsnGlnMetHis...HisAspValAsnProPheP 508

1687 .....GCAGTGGAGATTCCAGAAGAAGGCTTCTTATTGGGATGTGTGTTT 1731
          :::::::::      |||   :::|||   |||      |||
 508 roIleGlnMetAsnValAspProGluLeuTyrLeuThrGlyIleHisPhe 524

1732 GCAATTGCGGATTATCCAGAGCAGATGTCTGATAAGCAACTGCTGGCCAC 1781
           |||      :::   ::::::||||||             ::
 525 HisCysPheAspSerGluLysLeuPheGluAspLysLeuAspArgLysAs 541

1782 CTGGAAAAGGATAATCCAGGCACATGGCGGCACTGTTGAC.........C 1822
     :   :::   :::|||:::   ||||||||   ::::::
 541 nLeuGluPheMetIleLysTyrHisGlyGlyAspIleGluPheGluAlaS 558

1823 CCACCTTCACGAGTCGA.........TGCACGCACCTTCTCTGTGAGAGT 1863
        |||   :::|||           ||||||:::|||   :::|||
 558 erLysPheHisGluArgIleHisGlnIleThrHisValLeuValAspSer 574

1864 CAAGTCAGCAGCGCGTATGCACAGGCAATAAGAGAAAGAAAGAGATGTGT 1913
          ::::::           ::::::::::::|||||||||   ||
 575 CysArgAsnThrAsnValMetThrSerLeuGluHisArgLysArgIleVa 591

1914 TACTGCACACTGGTTAAACACAGTCTTAAAAAAAAAAAAAAATGGTACCGC 1963
     |:::   :::|||:::      |||:::::::::::::::::::      |
 591 lSerIleGlnTrpIleIleAspValMetGluArgGluArgValGluLeuP 608

1964 CGCACCGAGCCCTTCACTTCCCAGTGGCCTTCCCACCAGGAGGAAAGCCA 2013
     ||   |||      |||:::|||      |||      :::   :::|||
 608 roTrpArgIleAlaHisLeuProProProPheAsnGluAsnPheArgPro 624

2014 TGTTCACAGCATATTATTTCTGTGACTGGATTTGTTGATAGTGACAGAGA 2063
               ::::::|||:::::::||||||   :::|||:::|||
 625 TyrLeuGlyLysLeuPheSerLeuSerGlyPheAspSerSerGluArgAl 641

2064 TGACCTAAAATTAATGGCTTATTTGGCAGGTGCCAAATATACGGGTTATC 2113
     ::::::::::|||||||   :::   |||||||||   |||   :::|
 641 aAlaIleSerPheMetAlaGluMetMetGlyAlaLysIleThrProPheL 658

2114 TATGCCGCAGCAACACAGTCCTCATCTGTAAAGAACCAACTGGTTTAAAG 2163
     ||:::|||:::|||   :::|||||||:::|||:::|||:::      |||
 658 cuAlaArgGlnAsnAspLeuLeuIleAlaLysGlnProAlaAsp...Lys 673

2164 TATGAAAAAGCCAAAGAGTGGAGGATACCCTGTGTCAACGCCCAGTGGCT 2213
          :::|||:::||||||:::::|||   ||||||   '||||||::
 674 ValThrArgAlaGlnGluTrpLysValProValValAsnPheGlnTrpIl 690

2214 TGGCGACATTCTTCTGGGAAACTTTGAGGCACTGAGGCAG.......... 2253
     :::::|||      :::   :::   ||||||      |||
 690 eAlaAspAlaTyrValCysAsp...GluAlaSerProGlnLysArgProA 706

2254 ..ATTCAGTATAGTCGCTACACGGCATTCAGTCTGCAGGATCCATTTGCC 2301
       ::::::      ||||||          :::|||::::::|||
 706 snValGluAsnProArgTyr.........GlnLeuGluAsnProValMet 719

2302 CCTACCCAGCATTTAGTTTTAAAT........................... 2325
        :::   :::         |||
 720 ThrAsnValAsnThrSerProAsnHisIleGluGlyAlaSerGlnGluPh 736

2326 .......CTTTTAGATGCTTGGAGAGTTCCCTTAAAAGTGTCTGCAGAGT 2368
             :::::::::::::|||:::             ::::::   :::
 736 eGlyAlaMetMetSerCysTrpLysSerGlnThrIleIleAlaAspGlnA 753

2369 TGTTGATGAGTATAAGACTACCTCCCAAACTGAAACAGAATGAAGTA... 2415
          :::   |||         |||      ||||||::::::
 753 laTyrGluLysAlaArgGlnAsnArgLysThrLeuGlnAsnAsnThrPhe 769

2416 ..........................GCTAATGTCCAGCCTTCT..... 2433
                                  |||   |||:::
 770 PhePheProSerLysArgLeuSerAspGlnAlaValAlaProThrGluAs 786

2434 ..........TCCCAAAGAGCCAGAATTGAAGACGTA............. 2460
                |||:::|||||||||:::::::::::
 786 pGluIleLysSerSerArgAlaArgLeuAspGluIleLysThrHisTrpG 803

2461 ........................CCACCTCCCACTAAAAAGCTA... 2481
                                  |||   :::|||::::::
 803 luGluHisGlnGluLysMetLysAspThrProLeuSerLysGluPheTyr 819

2482 .........ACTCCAGAATTGACCCCTTTTGTGCTTTTC...ACTGGATT 2519
                   |||:::              :::   |||      |||::
 820 ArgValArgGluProGlnAlaArgValSerAlaTrpPheGlyGluGlyIl 836

2520 CGAGCCTGTCCAGGTTCAACAGTATATTAAGAAGCTCTACATTCTTGGTG 2569
     ::::          :::            |||||||||   :::|||||||
 836 eAspAspGluAlaLeuThrIleLeuLysLysLysLeuGluPheLeuGlyG 853

2570 GAGAGGTTGCGGAGTCTGCACAGAAGTGCACACACCTCATTGCCAGCAAA 2619
     |||||   :::|||:::   :::   :::|||||||:::|||      :::
 853 lyGluCysValGluLysIleArgAspAlaThrHisValIleLeuIleSer 869

2620 GTGACTCGCACCGTGAAGTTCCTGACGGCGATTTCTGTCGTGAAGCACAT 2669
          |||:::::::   :::|||   :::|||          |||:::|
 870 GlyArgArgSerLeuValLeuLeuGluSerIleIleArgGlyLysAsnIl 886

2670 AGTGACGCCAGAGTGGCTGGAAGAATGCTTCAGGTGTCAGAAGTTCATTG 2719
     |:::   ||||||||||:::   :::   ::::::   :::   ::::::|
 886 eMetAspProGluTrpIleValAspSerTyrLysGlnLysMetTrpLeuA 903

2720 ATGAGCAGAACTACATTCTCCGAGATGCTGAGGCAGAAGTACTTTTCTCT 2769
     ||      :::|||:::|||:::|||   |||   |||      |||:::
 903 spThrLeuAspTyrPheLeuHisAspLysGluLeuGluLysGluPheAla 919

2770 TTCAGCTTGGAAGAATCCTTAAAACGGGCACACGTTTCTCCACTCTTTAA 2819
     ::::::   :::::::|||:::   |||||||::   :::|||:::|||::
 920 TyrAsnCys ysArgSerValLeuArgAlaArgAsnLysProValPheGl 936

2820 GGCAAAATATTTTTACATCACACCTGGAATCTGCCCAAGTCTTTCCACTA 2869
     :         |||:::::::|||      :::   |||:::   :::   :
 936 uAspIleGluPheHisValThrArgPheValGluProAsnGlnAsnAspL 953

2870 TGAAGGCAATCGTAGAGTGTGCAGGAGGAAAGGTGTTATCCAAGCAGCCA 2919
     ::      ::::::|||   :::|||||||:::|||   |||:::::|||
 953 euValArgLeuIleGluLeuGlyGlyGlyAsnValHisSerGluLysPro 969

2920 TCT.........TTCCGGAAGCTCATGGAGCACAAGCAGAACTCGAGTTT 2960
     :::          :::   |||   :::|||   :::|||
 970 AspProLysTyrLeuAlaLysCysValGluThrGluGln........... 982

2961 GTCGGAAATAATTTTAATATCCTGTGAAAATGACCTTCATTTATGCCGAG 3010
          :::|||:::|||||||||||||||||||   ::::::
 983 ....ProPheIleIleIleSerCysGluAsnAspAlaArgPheLeuSerT 998

3011 AATATTTTGCCAGAGGCATAGATGTTCACAATGCAGAGTTCGTTCTGACT 3060
          :::   :::   :::::::|||:::::::::|||||||
 998 yrLeuAlaGlySerLysLeuProIleTyrAsnValAspLeuValLeuPhe 1014

3061 GGAGTGCTCACTCAAACGCTGGAC 3084
     ::::::|||   |||   ::::::
1015 AlaMetLeuArgGlnGlnIleGlu 1022
```

seq_name: sp_invertebrate:Q9U370

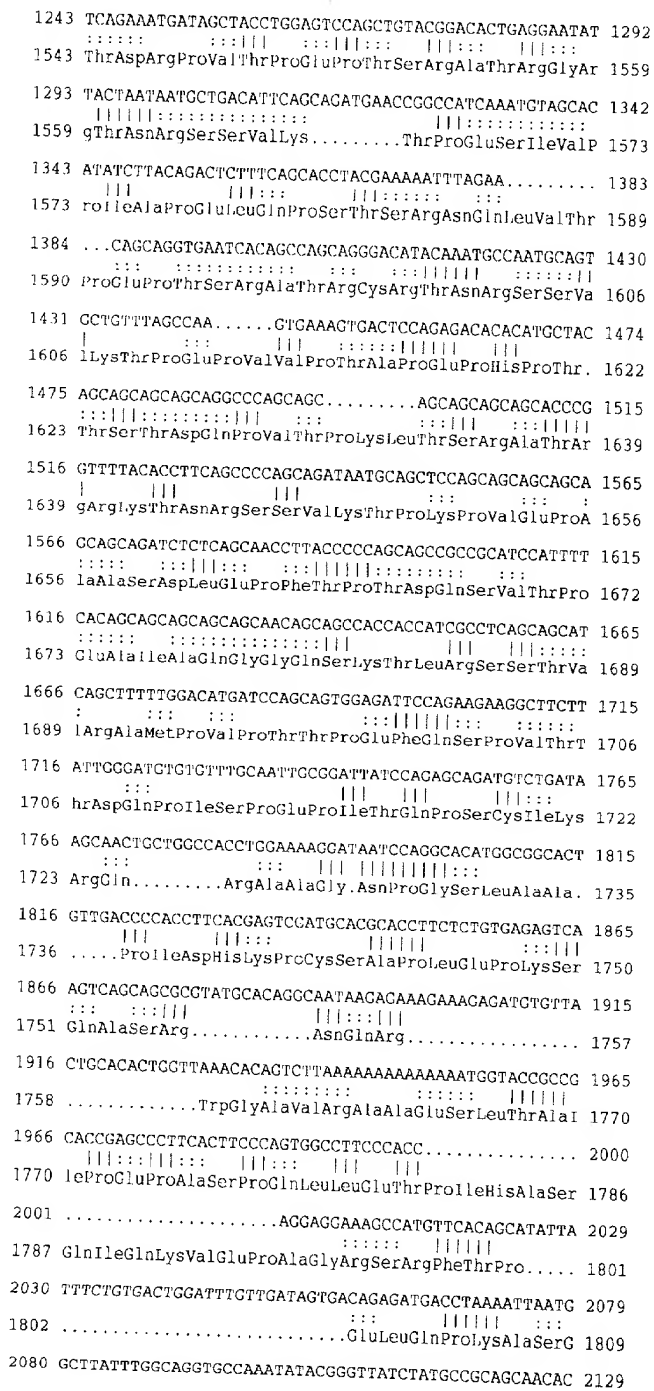
seq_documentation_block:

```
ID   Q9U370      PRELIMINARY;      PRT;  1076 AA.
AC   Q9U370;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   T13F2.3B PROTEIN.
GN   T13F2.3B.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
```

```
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Swinburne J.;
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; Z81122; CAB54313.1; -.
DR   InterPro; IPR001357; BRCT.
DR   InterPro; IPR002965; P_rich_extensn.
DR   Pfam; PF00533; BRCT; 4.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   SMART; SM00292; BRCT; 4.
DR   PROSITE; PS50172; BRCT; 4.
SQ   SEQUENCE   1076 AA;  120541 MW;  206A94A192802843 CRC64;


alignment_scores:
            Quality:  524.00              Length:    692
              Ratio:  1.444                Gaps:     24
 Percent Similarity:  52.457   Percent Identity: 25.000

alignment_block:
 US-09-664-641-10 x Q9U370   ..

 Align seq 1/1  to: Q9U370  from: 1  to: 1076

    1350 ACAGACTCTTTCAGCACC......TACGAAAAATTTAGAACAGCAGGTGA 1393
          |||   :::|||||||:::      ::::::   ||||||   :::
     349 ThrProThrPheSerSerProAlaHisGlnMetPheArgGlnGlyGlnPr 365

    1394 ATCACAGCCAGCAGGGACATACAAATGCCAATGCAGTGCTGTTTAGCCAA 1443
          :::   |||   |||      :::                :::   |||:
     365 oThrAlaProProGlyMetArgGln............TyrIleSerProG 378

    1444 GTGAAAGTGACTCCAGAGAGACACATGCTACAGCA.........GCAGCA 1484
          ::   ::::::::::::   |||   :::|||              |||
     378 lnAlaGlyGlnGlnGlnProThrProGlyThrProHisArgPheAlaPro 394

    1485 GCAGGCCCAGCAGCAGCAGCAGCAGCACCCGGTTTTACACCTTCAGCCCC 1534
          |||||||||||||||||||||||||||   |||:::|||:::||||||
     395 ProGlyProSerGlyThrThrSerSerProHisPheAlaProGlyAlaPr 411

    1535 AGCAGATAAT.................GCAGCTCCAGCAGCAG:AGCAG 1566
          |:::   :::                      :::||| :::|||||:::
     411 oGlyAlaAspLeuArgSerProSerLeuMetSerPro.AsnGlnGlnSer 427

    1567 CAGCAG..................................................... 1572
          ||||||
     428 GlnGlnHisThrProValLeuProArgSerLeuHisAsnGluMetThrSe 444

    1573 .....................ATCTCTCAGCAACCTTACCCCCAGCAG. 1599
                              :::   :::|||||   ||||||
     444 rHisGlyMetArgProIleAlaLeuProSerGlnProLeuProGlnThrG 461

    1600 .....................................CCGCCGCAT... 1608
                                             ||||||:::
     461 lnGlyThrProGlnProIleLeuSerAlaSerSerLeuProProGlnSer 477

    1609 ............CCATTTTCACAGCAGCAGCAGCAGCAACAGCAGCCACC 1646
                     |||      :::   ||||||   :::|||
     478 AlaGlyAlaThrGlnPheValTyrArgProGlnGlnProAsnGlnMetAs 494

    1647 ACCATCGCCTCAGCAGCATCAGCTTTTTGGACATGATCCA.......... 1686
                  |||::::::::|||:::      ||||||
     494 nThrMetAsnGlnMetAsnGlnMetHis...HisAspValAsnProPheP 510

    1687 .....GCAGTGGAGATTCCAGAAGAAGGCTTCTTATTGGGATGTGTGTTT 1731
                :::::::::      |||   :::|||   |||      |||
     510 roIleGlnMetAsnValAspProGluLeuTyrLeuThrGlyIleHisPhe 526

    1732 GCAATTGCGGATTATCCAGAGCAGATGTCTGATAAGCAACTGCTGGCCAC 1781
                   |||      :::   ::::::||||||            ::
     527 HisCysPheAspSerGluLysLeuPheGluAspLysLeuAspArgLysAs 543

    1782 CTGGAAAAGGATAATCCAGGCACATGGCGGCACTGTTGAC.........C 1822
          :   :::   :::|||:::   |||||||||   ::::::
     543 nLeuGluPheMetIleLysTyrHisGlyGlyAspIleGluPheGluAlaS 560

    1823 CCACCTTCACGAGTCGA.........TGCACGCACCTTCTCTGTGAGAGT 1863
               |||   :::|||              |||||:::|||   :::|||
     560 erLysPheHisGluArgIleHisGlnIleThrHisValLeuValAspSer 576

    1864 CAAGTCAGCAGCGCGTATGCACAGGCAATAAGACAAAGAAAGAGATGTGT 1913
               ::::::              ::::::::::|||||||||   ||
     577 CysArgAsnThrAsnValMetThrSerLeuGluHisArgLysArgIleVa 593

    1914 TACTGCACACTGGTTAAACACAGTCTTAAAAAAAAAAAAAAATGGTACCGC 1963
          |:::   :::|||:::      |||:::::::::::::::::::      |
     593 lSerIleGlnTrpIleIleAspValMetGluArgGluArgValGluLeuP 610

    1964 CGCACCGAGCCCTTCACTTCCCAGTGGCCTTCCCACCAGGAGGAAAGCCA 2013
          ||   |||      |||:::|||      |||      :::   :::|||
     610 roTrpArgIleAlaHisLeuProProProProPheAsnGluAsnPheArgPro 626

    2014 TGTTCACAGCATATTATTTCTGTGACTGGATTTGTTGATAGTGACAGAGA 2063
                   :::::::|||:::::|||||   :::|||:::|||
     627 TyrLeuGlyLysLeuPheSerLeuSerGlyPheAspSerSerGluArgAl 643

    2064 TGACCTAAAATTAATGGCTTATTTGGCAGGTGCCAAATATACGGGTTATC 2113
          :::::::::|||||   :::   |||||||||   |||   :::|
     643 aAlaIleSerPheMetAlaGluMetMetGlyAlaLysIleThrProPheL 660

    2114 TATGCCGCAGCAACACAGTCCTCATCTGTAAAGAACCAACTGGTTTAAAG 2163
          ||:::|||:::|||   :::||||||:::|||:::|||:::      |||
     660 euAlaArgGlnAsnAspLeuLeuIleAlaLysGlnProAlaAsp...Lys 675

    2164 TATGAAAAAGCCAAAGAGTGGAGGATACCCTGTGTCAACGCCCAGTGGCT 2213
          :::|||:::||||||:::::|||   ||||||   ||||||::
     676 ValThrArgAlaGlnGluTrpLysValProValValAsnPheGlnTrpIl 692

    2214 TGGCGACATTCTTCTGGGAAACTTTGAGGCACTGAGGCAG..... 2253
          :::|||      :::   :::   ||||||      |||
     692 eAlaAspAlaTyrValCysAsp...GluAlaSerProGlnLysArgProA 708

    2254 ..ATTCAGTATAGTCGCTACACGGCATTCAGTCTGCAGGATCCATTTGCC 2301
            ::::::      ||||||         :::|||:::::::|||
     708 snValGluAsnProArgTyr.........GlnLeuGluAsnProValMet 721

    2302 CCTACCCAGCATTTAGTTTTAAAT..... 2325
          :::   :::         |||
     722 ThrAsnValAsnThrSerProAsnHisIleGluGlyAlaSerGlnGluPh 738

    2326 .......CTTTTAGATGCTTGGAGAGTTCCCTTAAAAGTGTCTGCAGAGT 2368
                :::::::::::|||:::          ::::::   :::
     738 eGlyAlaMetMetSerCysTrpLysSerGlnThrIleIleAlaAspGlnA 755

    2369 TGTTGATGAGTATAAGACTACCTCCCAAACTGAAACAGAATGAAGTA... 2415
                :::   |||         |||      ||||||::::::
     755 laTyrGluLysAlaArgGlnAsnArgLysThrLeuGlnAsnAsnThrPhe 771

    2416 ..........................GCTAATGTCCAGCCTTCT..... 2433
                                    |||   |||:::
     772 PhePheProSerLysArgLeuSerAspGlnAlaValAlaProThrGluAs 788

    2434 ..........TCCCAAAGAGCCAGAATTGAAGACGTA..... 2460
                  |||:::||||||||||:::::::::::::
     788 pGluIleLysSerSerArgAlaArgLeuAspGluIleLysThrHisTrpG 805

    2461 ......... ..............CCACCTCCCACTAAAAAGCTA... 2481
                                  |||   :::|||::::::
```

```
 805 luGluHisGlnGluLysMetLysAspThrProLeuSerLysGluPheTyr 821

2482 .........ACTCCAGAATTGACCCCTTTTGTGCTTTTC...ACTGGATT 2519
                |||:::          :::   |||     |||::
 822 ArgValArgGluProGlnAlaArgValSerAlaTrpPheGlyGluGlyIl 838

2520 CGAGCCTGTCCAGGTTCAACAGTATATTAAGAAGCTCTACATTCTTGGTG 2569
     ::::         :::          |||||||||   :::|||||||
 838 eAspAspGluAlaLeuThrIleLeuLysLysLysLeuGluPheLeuGlyG 855

2570 CAGAGGTTGCGGAGTCTGCACAGAAGTGCACACACCTCATTGCCAGCAAA 2619
     |||||   :::|||:::   :::   :::|||||||:::|||      :::
 855 lyGluCysValGluLysIleArgAspAlaThrHisValIleLeuIleSer 871

2620 GTGACTCGCACCGTGAAGTTCCTGACGGCGATTTCTGTCGTGAAGCACAT 2669
          |||::::::   :::|||   :::|||         |||:::||
 872 GlyArgArgSerLeuValLeuLeuGluSerIleIleArgGlyLysAsnIl 888

2670 AGTGACGCCAGAGTGGCTGGAAGAATGCTTCAGGTGTCAGAAGTTCATTG 2719
     |:::   ||||||||||:::   :::   ::::::   :::   ::::::|
 888 eMetAspProGluTrpIleValAspSerTyrLysGlnLysMetTrpLeuA 905

2720 ATGAGCACAACTACATTCTCCGAGATGCTGAGGCAGAAGTACTTTTCTCT 2769
     ||      :::|||:::|||:::|||   |||   |||      |||:::
 905 spThrLeuAspTyrPheLeuHisAspLysGluLeuGluLysGluPheAla 921

2770 TTCAGCTTGGAAGAATCCTTAAAACGGGCACACGTTTCTCCACTCTTTAA 2819
     ::::::   ::::::|||:::   |||||||:::   :::|||:::|||:::
 922 TyrAsnCysLysArgSerValLeuArgAlaArgAsnLysProValPheGl 938

2820 GGCAAAATATTTTTACATCACACCTGGAATCTGCCCAAGTCTTTCCACTA 2869
     :         |||:::::::|||      :::   |||:::   :::   :
 938 uAspIleGluPheHisValThrArgPheValGluProAsnGlnAsnAspL 955

2870 TGAAGGCAATCGTAGAGTGTGCAGGAGGAAAGGTGTTATCCAAGCAGCCA 2919
     ::      ::::::|||   :::||||||:::|||   |||:::::::|||
 955 euValArgLeuIleGluLeuGlyGlyGlyAsnValHisSerGluLysPro 971

2920 TCT.........TTCCGGAAGCTCATGGAGCACAAGCAGAACTCGAGTTT 2960
     :::         :::   |||   :::|||   ::::||
 972 AspProLysTyrLeuAlaLysCysValGluThrGluGln........... 984

2961 GTCGGAAATAATTTTAATATCCTGTGAAAATGACCTTCATTTATGCCGAG 3010
            :::|||:::|||||||||||||||||||   ::::::
 985 ....ProPheIleIleIleSerCysGluAsnAspAlaArgPheLeuSerT 1000

3011 AATATTTTGCCAGAGGCATAGATGTTCACAATGCAGAGTTCGTTCTGACT 3060
            :::      :::   ::::::|||:::::::::||||||
1000 yrLeuAlaGlySerLysLeuProIleTyrAsnValAspLeuValLeuPhe 1016

3061 GGAGTGCTCACTCAAACGCTGGAC 3084
     ::::::|||   |||   ::::::
1017 AlaMetLeuArgGlnGlnIleGlu 1024
```

```
seq_name: sp_human:Q14676

seq_documentation_block:
ID   Q14676      PRELIMINARY;      PRT;  2089 AA.
AC   Q14676;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   KIAA0170 PROTEIN.
GN   KIAA0170.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=MYELOID LEUKEMIA CELLS;
RX   MEDLINE=96281124; PubMed=8724849;
RA   Nagase T., Seki N., Ishikawa K.-I., Tanaka A., Nomura N.;
RT   "Prediction of the coding sequences of unidentified human genes. V.
RT   The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by
RT   analysis of cDNA clones from human cell line KG-1.";
RL   DNA Res. 3:17-24(1996).
DR   EMBL; D79992; BAA11487.1; -.
DR   InterPro; IPR001357; BRCT.
DR   InterPro; IPR000253; FHA_domain.
DR   Pfam; PF00498; FHA; 1.
DR   SMART; SM00292; BRCT; 1.
DR   SMART; SM00240; FHA; 1.
DR   PROSITE; PS50172; BRCT; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
SQ   SEQUENCE   2089 AA;  226690 MW;  A628ECC7567EEE50 CRC64;


alignment_scores:
           Quality:  342.50              Length:    842
             Ratio:  0.839                Gaps:     33
 Percent Similarity:  48.456   Percent Identity: 22.922

alignment_block:
 US-09-664-641-10 x Q14676   ..

 Align seg 1/1  to: Q14676  from: 1  to: 2089

 688 ATTGTTCCAGAGCCAAAGGGGGGGAAATACGAATGTGCTTTAAAGCGAGC 737
     ::::::|||||||||   :::         :::         :::|||::
1383 ValThrProGluProThrSerArgAlaThrArgGlyArgLysAsnArgSe 1399

 738 AAGTATTAAAATTGTGACTCCTGACTGGGTTCTGGATTGCGTATCAGAGA 787
     :|||  |||      ||||||:::  ::::::     :::
1399 rSerGlyLys......ThrProGluThrLeuValProThrAla....... 1411

 788 AAACCAAAAAGGACGAAGCATTTTATCATCCTCGTCTGATTATTTATGAA 837
                                  |||:::|||
1412 ............................ProLysLeu............ 1414

 838 GAGGAAGAAGAGGAAGAGGAAGAGGAGGAGGAAGTAGAAAATGAGGAACA 887
           |||  :::  :::  ::::::  |||      |||
1415 ......GluProSerThrSerThrAspGlnProValThrProGluProTh 1429

 888 AGATTCTCAGAATGAGGGTAGTACAGATGAGAAGTCA.............. 924
      ::::::  ::::::|||  |||:::::::::|||
1429 rSerGlnAlaThrArgGlyArgThrAsnArgSerSerValLysThrProG 1446

 925 ...........AGCCCTGCCAGCTCTCAAGAAGGGTCTCCTTCAGGTGAC 963
                     |||  :::  :::  |||  |||  |||
1446 luThrValValProThrAlaProGluLeuGlnProSerThrSerThrAsp 1462

 964 CAGCAGTTTTCACCTAAATCCAACACTGAAAAATCTAAAGGGGAATTAAT 1013
     |||      :::|||:::  :::::::::  ::::::|||:::
1463 GlnProValThrProGluProThrSerGlnAlaThrArgGlyArgThr.. 1478

1014 GTTTGATGATTCTTCAGATTCATCACCGGAAAAACAGGAGAGAAATTTAA 1063
        |||   ||||||  ::::::||||||
1479 ....AspArgSerSerValLysThrProGluThr.............. 1488

1064 ACTGGACCCCGGCCGAAGTCCCACAGTTA...GCTGCAGCAAAACGCAGG 1110
           :::         :::|||:::|||  |||:::|||:::
1489 .....ValValProThrAlaProGluLeuGlnAlaSerAlaSerThrAsp 1503

1111 CTGCCTCAGGGAAAGGAGCCT.................GGGTTGATTAA 1142
        |||      :::||||||                 |||      ||
1504 GlnProValThrSerGluProThrSerArgThrThrArgGlyArgLysAs 1520

1143 CTTGTGTGCCAATGTCCCACCCGTCCCAGGTAACATTTTGCCCCCTGAGG 1192
     |                          :::::::::      ||||||
1520 n.......................ArgSerSerValLysThrProGlu. 1528

1193 TCCGGGGTAATTTAATGGCTGCTGGACAAAACCTCCAAAGTTCTGAAAGA 1242
               :::::::::   |||:::   :::|||||||
1529 ........ThrValValProAlaAlaProGluLeuGlnProProThrSer 1542
```


1243 TCAGAAATGATAGCTACCTGGAGTCCAGCTGTACGGACACTGAGGAATAT 1292
1543 ThrAspArgProValThrProGluProThrSerArgAlaThrArgGlyAr 1559
1293 TACTAATAATGCTGACATTCAGCAGATGAACCGGCCATCAAATGTAGCAC 1342
1559 qThrAsnArgSerSerValLys.....ThrProGluSerIleValP 1573
1343 ATATCTTACAGACTCTTTCAGCACCTACGAAAAATTTAGAA..... 1383
1573 roIleAlaProGluLeuGlnProSerThrSerArgAsnGlnLeuValThr 1589
1384 ...CAGCAGGTGAATCACAGCCAGCAGGGACATACAAATGCCAATGCAGT 1430
1590 ProGluProThrSerArgAlaThrArgCysArgThrAsnArgSerServa 1606
1431 GCTGTTTAGCCAA......GTGAAAGTGACTCCAGAGACACACATGCTAC 1474
1606 lLysThrProGluProValValProThrAlaProGluProHisProThr. 1622
1475 AGCAGCAGCAGCAGGCCCAGCAGC.....AGCAGCAGCAGCACCCG 1515
1623 ThrSerThrAspGlnProValThrProLysLeuThrSerArgAlaThrAr 1639
1516 GTTTTACACCTTCAGCCCCAGCAGATAATGCAGCTCCAGCAGCAGCAGCA 1565
1639 gArgLysThrAsnArgSerServalLysThrProLysProValGluProA 1656
1566 GCAGCAGATCTCTCAGCAACCTTACCCCCAGCAGCCGCCGCATCCATTTT 1615
1656 laAlaSerAspLeuGluProPheThrProThrAspGlnSerValThrPro 1672
1616 CACAGCAGCAGCAGCAGCAACAGCAGCCACCACCATCGCCTCAGCAGCAT 1665
1673 GluAlaIleAlaGlnGlyGlyGlnSerLysThrLeuArgSerSerThrVa 1689
1666 CAGCTTTTTGGACATGATCCAGCAGTGGAGATTCCAGAAGAAGGCTTCTT 1715
1689 lArgAlaMetProValProThrThrProGluPheGlnSerProValThrT 1706
1716 ATTGGGATGTGTGTTTGCAATTGCGGATTATCCAGAGCAGATGTCTGATA 1765
1706 hrAspGlnProIleSerProGluProIleThrGlnProSerCysIleLys 1722
1766 AGCAACTGCTGGCCACCTGGAAAAGGATAATCCAGGCACATGGCGGCACT 1815
1723 ArgGln.....ArgAlaAlaGly.AsnProGlySerLeuAlaAla. 1735
1816 GTTGACCCCACCTTCACGAGTCCATGCACGCACCTTCTCTGTGAGAGTCA 1865
1736 .....ProIleAspHisLysProCysSerAlaProLeuGluProLysSer 1750
1866 AGTCAGCAGCGCGTATGCACAGGCAATAAGACAAAGAAAGAGATGTGTTA 1915
1751 GlnAlaSerArg.....AsnGlnArg..... 1757
1916 CTGCACACTGGTTAAACACAGTCTTAAAAAAAAAAAAAAATGGTACCGCCG 1965
1758 .....TrpGlyAlaValArgAlaAlaGluSerLeuThrAlaI 1770
1966 CACCGAGCCCTTCACTTCCCAGTGGCCTTCCCACC..... 2000
1770 leProGluProAlaSerProGlnLeuLeuGluThrProIleHisAlaSer 1786
2001 .....AGGAGGAAAGCCATGTTCACAGCATATTA 2029
1787 GlnIleGlnLysValGluProAlaGlyArgSerArgPheThrPro..... 1801
2030 TTTCTGTGACTGGATTTGTTGATAGTGACAGAGATGACCTAAAATTAATG 2079
1802 .....GluLeuGlnProLysAlaSerG 1809
2080 GCTTATTTGGCAGGTGCCAAATATACGGGTTATCTATGCCGCAGCAACAC 2129
1809 lnSerArgLysArgSerLeuAlaThrMetAspSerProProHisGlnLys 1825
2130 AGTCCTCATCTGTAAAGAACCAACTGGTTTAAAGTATGAAAAAGCCAAAG 2179
1826 GlnProGlnArgGlyGluValSerGlnLysThrValIleIleLysGluGl 1842
2180 AGTGGAGGATACCCTGTGTCAACGCCCAGTGGCTTGGCGACATTCTTCTG 2229
1842 uGluGluAspThr...AlaGluLysPro..... 1850
2230 GGAAACTTTGAGGCACTGAGGCAGATTCAGTATAGTCGCTACACGGCATT 2279
1851 .....GlyLysGluGluAspValVal..... 1857
2280 CAGTCTGCAGGATCCATTTGCCCCTACCCAGCATTTAGTTTTAAATCTTT 2329
1858 .....ThrProLysProGly..... 1862
2330 TAGATGCTTGGAGAGTTCCCTTAAAAGTGTCTGCAGAGTTGTTGATGAGT 2379
1863 .....L 1863
2380 ATAAGACTACCTCCCAAACTGAAACAGAATGAAGTAGCTAATGTCCAGCC 2429
1863 ysArgLysArgAspGlnAlaGluGluGlu.....ProAsn 1874
2430 TTCTTCCCAAAGAGCCAGAATTGAAGACGTACCACCTCCCACTAAAAAGC 2479
1875 ArgIlePro...SerArgSerLeuArgArgThr.....LysLe 1886
2480 TAACTCCAGAATTGACC...CCTTTTGTGCTTTTCACTGGATTCGAGCCT 2526
1886 uAsnGln.GluSerThrAlaProLysValLeuPheThrGlyValValAsp 1902
2527 GTCCAGGTTCAACAGTATATTAAGAAGCTCTACATTCTTGGTGGAGAGGT 2576
1903 AlaArgGlyGlu.....ArgAlaValLeuAlaLeuGlyGlySerLe 1916
2577 TGCGGAGTCTGCACAGAAGTGCACACACCTCATTGCCAGCAAAGTGACTC 2626
1916 uAlaGlySerAlaAlaGluAlaSerHisLeuValThrAspArgIleArgA 1933
2627 GCACCGTGAAGTTCCTGACGGCGATTTCTGTCGTGAAGCACATAGTGACG 2676
1933 rgThrValLysPheLeuCysAlaLeuGlyArgGlyIleProIleLeuSer 1949
2677 CCAGAGTGGCTGGAAGAATGCTTCAGGTGTCAGAAGTTCATTGATGAGCA 2726
1950 LeuAspTrpLeuHisGlnSerArgLysAlaGlyPhePheLeuProProAs 1966
2727 GAACTACATTCTCCGAGATGCTGAGGCAGAAGTACTTTTCTCTTTCAGCT 2776
1966 pGluTyrValValThrAspProGluGlnGluLysAsnPheGlyPheSerL 1983
2777 TGGAAGAATCCTTAAAACGGGCACACGTTTCTCCACTCTTTAAGGCAAAA 2826
1983 euGlnAspAlaLeuSerArgAlaArgGluArgArgLeuLeuGluGlyTyr 1999
2827 TATTTTTACATCACACCTGGAATCTGCCCAAGTCTTTCCACTATGAAGGC 2876
2000 GluIleTyrValThrProGlyValGlnProProProProGlnMetGlyGl 2016
2877 AATCGTAGAGTGTGCAGGAGGAAAGGTGTTATCCAAGCAGCCATCTTTCC 2926
2016 uIleIleSerCysCysGlyGlyThrTyrLeuProSerMetPro..... 2030
2927 GGAAGCTCATGGAGCACAAGCAGAACTCGAGTTTGTCGGAAATAATTTTA 2976
2031 .....ArgSerTyrLysProGlnArgValVal 2039
2977 ATATCCTGTGAAAATGACCTTCATTTATGCCGAGAATATTTTGCCAGAGG 3026

```
 2040 IleThrCysProGlnAspPheProHisCysSerIleProLeuArgValGl 2056

 3027 CATAGATGTTCACAATGCAGAGTTCGTTCTGACTGGAGTGCTCACTCAAA 3076
      |:::   :::   :::   ||||||:::|||||||||||||||   |||
 2056 yLeuProLeuLeuSerProGluPheLeuLeuThrGlyValLeuLysGlnG 2073

 3077 CGCTGGACTATGAATCATAT 3096
                 |||::::::
 2073 luAlaLysProGluAlaPhe 2079

seq_name: sp_human:Q96QC2

seq_documentation_block:
ID   Q96QC2      PRELIMINARY;      PRT;  2090 AA.
AC   Q96QC2;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   KIAA0170 PROTEIN.
GN   KIAA0170.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Hirakawa M., Yamaguchi H., Imai K., Shimada J.;
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Shiina S., Tamiya G., Oka A., Inoko H.;
RT   "Homo sapiens 2,229,817bp genomic DNA of 6p21.3 HLA class I region.";
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP000512; BAB63322.1; -.
SQ   SEQUENCE   2090 AA;  226732 MW;  DEB2CECB9A0711AC CRC64;


alignment_scores:
           Quality:  340.50              Length:    842
             Ratio:  0.833                Gaps:     33
 Percent Similarity:  48.575   Percent Identity: 22.803

alignment_block:
 US-09-664-641-10 x Q96QC2   ..

 Align seg 1/1  to: Q96QC2  from: 1  to: 2090

  688 ATTGTTCCAGAGCCAAAGGGGGGGAAATACGAATGTGCTTTAAAGCGAGC 737
      ::::::||||||||||  :::        :::        :::|||::
 1384 ValThrProGluProThrSerArgAlaThrArgGlyArgLysAsnArgSe 1400

  738 AAGTATTAAAATTGTGACTCCTGACTGGGTTCTGGATTGCGTATCAGAGA 787
      :|||  |||      |||||||:::  ::::::     :::
 1400 rSerGlyLys......ThrProGluThrLeuValProThrAla..... 1412

  788 AAACCAAAAAGGACGAAGCATTTTATCATCCTCGTCTGATTATTTATGAA 837
                                |||:::|||
 1413 ..........................ProLysLeu............ 1415

  838 GAGGAAGAAGAGGAAGAGGAAGAGGAGGAGGAAGTAGAAAATGAGGAACA 887
          |||  :::  :::  ::::::  |||     |||
 1416 ......GluProSerThrSerThrAspGlnProValThrProGluProTh 1430

  888 AGATTCTCAGAATGAGGGTAGTACAGATGAGAAGTCA.............. 924
      ::::::  ::::::|||  |||:::::::::|||
 1430 rSerGlnAlaThrArgGlyArgThrAsnArgSerSerValLysThrProG 1447

  925 ..........AGCCCTGCCAGCTCTCAAGAAGGGTCTCCTTCAGGTGAC 963
                 |||  :::  :::  |||  |||  |||
 1447 luThrValValProThrAlaProGluLeuGlnProSerThrSerThrAsp 1463

  964 CAGCAGTTTTCACCTAAATCCAACACTGAAAAATCTAAAGGGGAATTAAT 1013
      |||      :::|||:::  :::::::::  ::::::|||:::

 1464 GlnProValThrProGluProThrSerGlnAlaThrArgGlyArgThr.. 1479

 1014 GTTTGATGATTCTTCAGATTCATCACCGGAAAAACAGGAGAGAAATTTAA 1063
           |||   ||||||   ::::::::||||||
 1480 ....AspArgSerSerValLysThrProGluThr.............. 1489

 1064 ACTGGACCCCGGCCGAAGTCCCACAGTTA...GCTGCAGCAAAACGCAGG 1110
           :::          :::|||:::|||   |||:::|||:::
 1490 .....ValValProThrAlaProGluLeuGlnAlaSerAlaSerThrAsp 1504

 1111 CTGCCTCAGGGAAAGGAGCCT.................GGGTTGATTAA 1142
           |||       :::::::|||                |||      ||
 1505 GlnProValThrSerAspProThrSerArgThrThrArgGlyArgLysAs 1521

 1143 CTTGTGTGCCAATGTCCCACCCGTCCCAGGTAACATTTTGCCCCCTGAGG 1192
      |                                 :::::::::      ||||||
 1521 n.......................ArgSerSerValLysThrProGlu. 1529

 1193 TCCGGGGTAATTTAATGGCTGCTGGACAAAACCTCCAAAGTTCTGAAAGA 1242
           :::::::::   |||:::   :::||||||
 1530 ........ThrValValProAlaAlaProGluLeuGlnProProThrSer 1543

 1243 TCAGAAATGATAGCTACCTGGAGTCCAGCTGTACCGACACTGAGGAATAT 1292
      :::::::::   :::|||   :::|||:::   |||:::   |||:::
 1544 ThrAspGlnProValThrProGluProThrSerArgAlaThrArgGlyAr 1560

 1293 TACTAATAATGCTGACATTCAGCAGATGAACCGGCCATCAAATGTAGCAC 1342
      ||||||:::::::::::::::                ||||:::::::::::
 1560 gThrAsnArgSerSerValLys.........ThrProGluSerIleValP 1574

 1343 ATATCTTACAGACTCTTTCAGCACCTACGAAAAATTTAGAA......... 1383
      |||         |||:::      |||:::::::   :::
 1574 roIleAlaProGluLeuGlnProSerThrSerArgAsnGlnLeuValThr 1590

 1384 ...CAGCAGGTGAATCACAGCCAGCAGGGACATACAAATGCCAATGCAGT 1430
           :::  ::::::::::::   :::   :::|||||||   :::::::||
 1591 ProGluProThrSerArgAlaThrArgCysArgThrAsnArgSerSerVa 1607

 1431 GCTGTTTAGCCAA......GTGAAAGTGACTCCAGAGACACACATGCTAC 1474
      |          :::      |||   ::::::||||||   |||
 1607 lLysThrProGluProValValProThrAlaProGluProHisProThr. 1623

 1475 AGCAGCAGCAGCAGGCCCAGCAGC.........AGCAGCAGCAGCACCCG 1515
      :::|||:::::::::|||   :::          :::|||   :::||||||
 1624 ThrSerThrAspGlnProValThrProLysLeuThrSerArgAlaThrAr 1640

 1516 GTTTTACACCTTCAGCCCCAGCAGATAATGCAGCTCCAGCAGCAGCAGCA 1565
      |     |||          |||              :::      :::    :
 1640 gArgLysThrAsnArgSerSerValLysThrProLysProValGluProA 1657

 1566 GCAGCAGATCTCTCAGCAACCTTACCCCCAGCAGCCGCCGCATCCATTTT 1615
      :::::   :::|||:::   :::||||||:::::::::   :::
 1657 laAlaSerAspLeuGluProPheThrProThrAspGlnSerValThrPro 1673

 1616 CACAGCAGCAGCAGCAGCAACAGCAGCCACCACCATCGCCTCAGCAGCAT 1665
      ::::::   ::::::::::::::::|||          |||   |||:::::
 1674 GluAlaIleAlaGlnGlyGlyGlnSerLysThrLeuArgSerSerThrVa 1690

 1666 CAGCTTTTTGGACATGATCCAGCAGTGGAGATTCCAGAAGAAGGCTTCTT 1715
      :      :::  :::              :::||||||:::   ::::::
 1690 lArgAlaMetProValProThrThrProGluPheGlnSerProValThrT 1707

 1716 ATTGGGATGTGTGTTTGCAATTGCGGATTATCCAGAGCAGATGTCTGATA 1765
           :::                |||  |||      |||:::
 1707 hrAspGlnProIleSerProGluProIleThrGlnProSerCysIleLys 1723

 1766 AGCAACTGCTGGCCACCTGGAAAAGGATAATCCAGGCACATGGCGGCACT 1815
      :::                :::  ||| ||||||||||:::
 1724 ArgGln.........ArgAlaAlaGly.AsnProGlySerLeuAlaAla. 1736

 1816 GTTGACCCCACCTTCACGAGTCGATGCACGCACCTTCTCTGTGAGAGTCA 1865
           |||      |||:::       ||||||          :::|||
 1737 .....ProIleAspHisLysProCysSerAlaProLeuGluProLysSer 1751
```

```
1866 AGTCAGCAGCGCGTATGCACAGGCAATAAGAGAAAGAAAGAGATGTGTTA 1915
1752 GlnAlaSerArg.............AsnGlnArg................. 1758

1916 CTGCACACTGGTTAAACACAGTCTTAAAAAAAAAAAAAATGGTACCGCCG 1965
1759 ..............TrpGlyAlaValArgAlaAlaGluSerLeuThrAlaI 1771

1966 CACCGAGCCCTTCACTTCCCAGTGGCCTTCCCACC................ 2000
1771 leProGluProAlaSerProGlnLeuLeuGluThrProIleHisAlaSer 1787

2001 ....................AGGAGCAAAGCCATGTTCACAGCATATTA 2029
1788 GlnIleGlnLysValGluProAlaGlyArgSerArgPheThrPro..... 1802

2030 TTTCTGTGACTGGATTTGTTGATACTGACAGAGATGACCTAAAATTAATG 2079
1803 ..........................GluLeuGlnProLysAlaSerG 1810

2080 GCTTATTTGGCAGGTGCCAAATATACGGGTTATCTATGCCGCAGCAACAC 2129
1810 lnSerArgLysArgSerLeuAlaThrMetAspSerProProHisGlnLys 1826

2130 AGTCCTCATCTGTAAAGAACCAACTGGTTTAAAGTATGAAAAAGCCAAAG 2179
1827 GlnProGlnArgGlyGluValSerGlnLysThrValIleIleLysGluGl 1843

2180 AGTGGAGGATACCCTGTGTCAACGCCCAGTGGCTTGGCGACATTCTTCTG 2229
1843 uGluGluAspThr...AlaGluLysPro..................... 1851

2230 GGAAACTTTGAGGCACTGAGGCAGATTCAGTATAGTCGCTACACGGCATT 2279
1852 ..........GlyLysGluGluAspValVal................... 1858

2280 CAGTCTGCAGGATCCATTTGCCCCTACCCAGCATTTAGTTTTAAATCTTT 2329
1859 ................ThrProLysProGly................... 1863

2330 TAGATGCTTGGAGAGTTCCCTTAAAAGTGTCTGCAGAGTTGTTGATGAGT 2379

1864 .................................................L 1864

2380 ATAAGACTACCTCCCAAACTGAAACAGAATGAAGTAGCTAATGTCCAGCC 2429
1864 ysArgLysArgAspGlnAlaGluGluGlu..............ProAsn 1875

2430 TTCTTCCCAAAGAGCCAGAATTGAAGACGTACCACCTCCCACTAAAAAGC 2479
1876 ArgIlePro...SerArgSerLeuArgArgThr............LysLe 1887

2480 TAACTCCAGAATTGACC...CCTTTTGTGCTTTTCACTGGATTCGAGCCT 2526
1887 uAsnGln.GluSerThrAlaProLysValLeuPheThrGlyValValAsp 1903

2527 GTCCAGGTTCAACAGTATATTAAGAAGCTCTACATTCTTGGTGGAGAGGT 2576
1904 AlaArgGlyGlu.........ArgAlaValLeuAlaLeuGlyGlySerLe 1917

2577 TGCGGAGTCTGCACAGAAGTGCACACACCTCATTGCCAGCAAAGTGACTC 2626
1917 uAlaGlySerAlaAlaGluAlaSerHisLeuValThrAspArgIleArgA 1934

2627 GCACCGTGAAGTTCCTGACGGCGATTTCTGTCGTGAAGCACATAGTGACG 2676
1934 rgThrValLysPheLeuCysAlaLeuGlyArgGlyIleProIleLeuSer 1950

2677 CCAGAGTGGCTGGAAGAATGCTTCAGGTGTCAGAAGTTCATTGATGAGCA 2726
1951 LeuAspTrpLeuHisGlnSerArgLysAlaGlyPhePheLeuProProAs 1967

2727 GAACTACATTCTCCGAGATGCTGAGGCAGAAGTACTTTTCTCTTTCAGCT 2776
1967 pGluTyrValValThrAspProGluGlnGluLysAsnPheGlyPheSerL 1984

2777 TGGAAGAATCCTTAAAACGGGCACACGTTTCTCCACTCTTTAAGGCAAAA 2826
1984 euGlnAspAlaLeuSerArgAlaArgGluArgArgLeuLeuGluGlyTyr 2000

2827 TATTTTTACATCACACCTGGAATCTGCCCAAGTCTTTCCACTATGAAGGC 2876
2001 GluIleTyrValThrProGlyValGlnProProProProGlnMetGlyGl 2017

2877 AATCGTAGAGTGTGCAGGAGGAAAGGTGTTATCCAAGCAGCCATCTTTCC 2926
2017 uIleIleSerCysCysGlyGlyThrTyrLeuProSerMetPro....... 2031

2927 GGAAGCTCATGGAGCACAAGCAGAACTCGAGTTTGTCGGAAATAATTTTA 2976
2032 .....................ArgSerTyrLysProGlnArgValVal 2040

2977 ATATCCTGTGAAAATGACCTTCATTTATGCCGAGAATATTTTGCCAGAGG 3026
2041 IleThrCysProGlnAspPheProHisCysSerIleProLeuArgValGl 2057

3027 CATAGATGTTCACAATGCAGAGTTCGTTCTGACTGGAGTGCTCACTCAAA 3076
2057 yLeuProLeuLeuSerProGluPheLeuLeuThrGlyValLeuLysGlnG 2074

3077 CGCTGGACTATGAATCATAT 3096
2074 luAlaLysProGluAlaPhe 2080
```

seq_name: sp_plant:Q9LVF7

seq_documentation_block:

```
ID   Q9LVF7       PRELIMINARY;      PRT;  1041 AA.
AC   Q9LVF7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   GB|AAD14441.1.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RA   Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RX   MEDLINE=20277480; PubMed=10819329;
RA   Nakamura Y.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT   features of the regions of 4,504,864 bp covered by sixty P1 and TAC
RT   clones.";
RL   DNA Res. 7:131-135(2000).
DR   EMBL; AB019232; BAB02343.1; -.
DR   InterPro; IPR001357; BRCT.
DR   Pfam; PF00533; BRCT; 1.
DR   SMART; SM00292; BRCT; 1.
DR   PROSITE; PS50172; BRCT; 1.
SQ   SEQUENCE   1041 AA;  115950 MW;  69B1C8B02A0374F4 CRC64;
```

alignment_scores:

```
            Quality:  316.50          Length:   1049
              Ratio:  0.745             Gaps:     44
Percent Similarity:  40.515   Percent Identity: 19.733
```

```
alignment_block:
 US-09-664-641-10 x Q9LVF7   ..

 Align seq 1/1  to: Q9LVF7  from: 1  to: 1041

    183 GCACAGGTTATTCAGCTTCTCAAGGCTGGAAAA......GCGAAGGAAGT 226
        ::::::   :::|||:::::|||:::   :::      ||||||:::||
    330 SerLysPheValGlnIlePheLysGlyProGlnGluLeuAlaLysLysVa 346

    227 TTCCTAC......AATGCACTAGCCTCACACATAATCTCAGAGGATGGGG 270
        |||||||       :::|||::::::::::::|||:::::::   |||   :
    346 lSerTyrLysSerGlyAlaValGlyAsnAsnIlePheAspTrpAspAspA 363

    271 ACAATCCAGAGGTGGGA.............................. 287
        :::::   :::   |||
    363 snArgGluAspGluGlyGlyGlyAspIleTyrArgArgArgLysAspGlu 379

    288 ..................GAAGCTCGGGAAGTCTTTGACTTACCTGTTGT 319
                         :::   ||||||      :::|||||||
    380 PhePheGlyValAlaSerLysArgArgGluPheSerSerLeuPro..... 394

    320 AAAGCCTTCTTGGGTGATTCTGTCCGTTCAGTGTGGAACTCTTCTGCCAG 369
                                 |||         |||:::|||||
    395 .....................ArgGluGlnLysArgGluLeuIleProV 404

    370 TAAATGGTTTTTCTCCAGAATCATGTCAGATTTTTTTTGGAATCACTGCC 419
        ||
    404 al.......................................... 404

    420 TGCCTTTCTCAGGGTGTTGATACAAGCTGG............AGCTCTTT 457
                  :::||||||      |||             |||:::||
    405 ............AlaValAspLysArgTrpAlaArgSerAspSerLysLe 417

    458 GTTGGAGTCTTCCAGAGCTCTCCCAAGGAGAGGTAGGGAAGGG....... 500
        ||||:::   |||   :::      |||::::::   :::|||
    417 uLeuLysHisSerValThrArgSerArgLysAsnIleGlnGlyAlaLysL 434

    501 ..AGCTTGTCCAGCAGAAGTTGGGAAGCACAGAGATCATCTGCCTTCTTC 548
          :::|||:::::: |||:::   :::   |||:::             ::
    434 ysAsnLeuGlyLys.GluLeuAspGluValArgGluAlaAlaValLeuGl 450

    549 TGACCCGGTATTGATGCAGGCTGAGGCCTCTGTTGTAATGTGCTGGGTGT 598
        ::::   :::   :::   |||||||||        :::|||
    450 yAsnAspThrGlnValAlaAlaGluAlaIleAspAspLeuCys....... 464

    599 CATCTGAAGACAGAAGTGCCCTGTGGGCTTTGGTTACGTTCTATGGGGGA 648
          |||   ||||||:::                          :::   |||
    465 ..SerGlyAspArgGlyLys.....................PheAspGly 473

    649 GATTGCCAGCTAACCCTCAATAAGAAATGCACGCATTTGATTGT.CCAGA 698
        :::::::::        :::||||||                :::   |||||
    474 GluAlaSerCysLeuThrGlyLysLys............LeuSerProGl 486

    699 GCCAAAGGGGGGGAAATACGAATGTGCTTTAAAGCGAGCAAGTATTAAAA 748
        |   :::   |||
    486 uGluGluArgGly................................... 490

    749 TTGTGACTCCTGACTGGGTTCTGGATTGCGTATCAGAGAAAACCAAAAAG 798
            :::|||      |||:::        |||:::   ||||||:::
    491 ..PheSerProGlyGlyValValThrArgGlnSerLysGlyThrLysArg 506

    799 GACGAAGCATTTTATCATCCTCGTCTGATTATTTATGAAGAGGAAGAAGA 848
        :::|||:::
    507 IleGlnAlaMet..................................... 510

    849 GGAAGAGGAAGAGGAGGAGGAAGTAGAAAATGAGGAACAAGATTCTCAGA 898
                :::::::::|||:::   :::::::::::
    511 .........SerLysAspGluLeuLeuLysLysArgMet.......... 520

    899 ATGAGGGTAGTACAGATGAGAAGTCAAGCCTTCCAGCTCTCAA...... 942
                        :::|||:::|||||:::   ::::::
    521 ................LysLysAlaSerProSerProAlaLysAlaCys 531

    943 ...........GAAGGGTCTCCTTCAGGTGACCAG.............. 966
                    |||||||||   :::||||||||||
    532 ArgThrAsnIleGluGlySerSerAsnGlyAspGlnLeuAsnLysGluGl 548

    966 ..................................................... 966

    548 yProCysCysTrpLysSerArgLysValGlnThrAlaSerArgGluThrL 565

    967 .........................CAGTTTTCACCTAAATCCAACACT 990
                                 :::   |||   :::|||||||||
    565 ysLysAsnLeuValAspGluPheAspGluValSerGlnGluSerAsnThr 581

    991 GAAAAATCTAAAGGGGAATTAATGTTTGAT......GATTCTTCAGATTC 1034
        |||                 |||||||||      ::::::::::::::
    582 Glu.................MetPheAspArgHisGluGluAlaGluAl 592

   1035 ATCACCGGAAAAACAGGAGAGAAATTTAAACTGGACCCCGGCCGAAGTCC 1084
        ::::|||:::   |||                          |||||||||
    592 aGlyProAspThrGln.................MetAlaAlaGluValM 603

   1085 CACAGTTAGCTGCAGCAAAACGCAGGCTGCCTCAGGGAAAGGAGCCTGGG 1134
          :::       :::               :::|||:::|||
    603 etAsnAlaLeuHisSer............GlyAspGlyArgGlu...... 613

   1135 TTGATTAACTTGTGTGCCAATGTCCCACCCGTCCCAGGTAACATTTTGCC 1184
                           :::   |||   |||:::|||::::::
    614 .....................IleAspProGluProAsnAsnLeuIleGl 623

   1185 CCCTGAG......GTCCGGGGTAATTTAATGGCTGCTGGACAAAACCTCC 1228
          :::      ::::::|||:::::      :::|||      :
    623 yLysLysLeuLeuLeuGluGlyGlyIleSerArgCysGlyValValThrA 640

   1229 AAAGTTCTGAAAGATCAGAAATGATAGCTACCTGGAGTCCAGCTGTACGG 1278
        :::::|||:::|||                              ::::::
    640 rgLysSerLysArg.............................IleLys 646

   1279 ACACTGAGGAATATTACTAATAATGCTGACATTCAGCAGATGAACCGGCC 1328
          ::::::   :::   |||       |||:::::::::::::   :::||
    647 GlyIleGlnAlaValAspAsn......AspValGluSerLeu...LysPr 660

   1329 ATCAAATGTAGCACATATCTTACAGACTCTTTCAGCACCTACGAAAAATT 1378
        |:::|||                  ::::::   |||      :::|||::::
    660 oLysAsn.............LysLysAlaArgSerIleLeuAlaLysSerP 673

   1379 TAGAACAGCAGGTGAATCACAGCCAGCAGGGACATACAAATGCCAATGCA 1428
        ::|||:::::::::::::::   :::::::::
    673 heGluLysAsnMetAspArgTyrSerLysAsnAsp.............. 684

   1429 GTGCTGTTTAGCCAAGTGAAAGTG...ACTCCAGAGACACACATGCTACA 1475
                        ||||||   ||||||:::      :::   ::
    685 .................LysValAspThrProAspGluAlaValAlaSe 695

   1476 GCAGCAGCAGCAGGCCCAG...CAGCAGCAGCAGCAGCACCCGGTTTTAC 1522
        :     ::::::   |||   :::   :::::::::|||   :::
    695 rThrThrGluLysArgGlnGlyGluLeuSerAsnLysHisCysMetSerL 712

   1523 ACCTTCAGCCCCAGCAGATAATGCAGCTCCAGCAGCAGCAGCAGCAGCAG 1572
          |||                     |||:::|||:::::::::   :::
    712 ysLeu.....................LeuLysGlnSerHisArgGlyGlu 721

   1573 ATCTCTCAGCAACCTTACCCCCAGCAGCCGCCGCATCCATTTTCACAGCA 1622
          :::         ||||||::::::                    :::||
    722 AlaGluValLeuAsnTyrProLysArgArgArgSerAlaArgIleSerGl 738

   1623 GCAGCAGCAGCAACAGCAGCCACCACCATCGCCTCAGCAGCATCAGCTTT 1672
        |:::|||   ::::::          |||
    738 nAspGlnValAsnGluAlaGlyArgSerSer................... 748

   1673 TTGGACATGATCCAGCAGTGGAGATTCCAGAAGAAGGCTTCTTATTGGGA 1722
              |||||||||   :::   |||   ::::::      :::
    749 ........AspProAlaPheAspThrProAlaLysSerLysThrProSer 762

   1723 TGTGTGTTTG AATTGCGGATTATCCAGAGCAGATGTCTGATAAGCAACT 1772
```

```
                :::             |||      |||:::|||
 763 ThrAsnValSer.............ProIleCysMetGlyAsp..... 772

1773 GCTGGCCACCTGGAAAAGGATAATCCAGGCACATGGCGGCACTGTTGACC 1822
                                :::   |||      :::
 773 ........................GluTyrHisArgLeuSerCysLysA 781

1823 CCACCTTCACGAGTCGATGCACCCACCTTCTCTGTGAGAGTCAAGTCAGC 1872
     :::||||||||||:::   |||
 781 spSerPheThrSerHisThrThr........................... 788

1873 AGCGCGTATGCACAGGCAATAAGAGAAAGAAAGAGATGTGTTACTGCACA 1922
                          ||||||
 789 ......................ArgGlu....................... 790

1923 CTGGTTAAACACAGTCTTAAAAAAAAAAAAAAATGGTACCGCCGCACCGAG 1972
                                                     |||:
 791 ..........................................PheArgS 793

1973 CCCTTCACTTCCCAGTGGCCTTCCCACCAGGAGGAAAGCCATGTTCACAG 2022
     ::|||      |||||||||   |||
 793 erLeuThrValProValAlaGluPro....................... 801

2023 CATATTATTTCTGTGACTGGATTTGTTGATAGTGACAGAGATGACCTAAA 2072
          ||||||   |||      :::   :::   ||||||
 802 ......IleSerGluThrLysSerThrArgLysArgArgAsp........ 813

2073 ATTAATGGCTTATTTGGCAGGTGCCAAATATACGGGTTATCTATGCCGCA 2122
                                      |||   :::|||
 814 ..............................LeuGlySerIleCys.... 818

2123 GCAACACAGTCCTCATCTGTAAAGAACCAACTGGTTTAAAGTATGAAAAA 2172
           ||||||:::
 819 ........ValLeuPhe................................. 821

2173 GCCAAAGAGTGGAGGATACCCTGTGTCAACGCCCAGTGGCTTGGCGACAT 2222

 821 ..................................................... 821

2223 TCTTCTGGGAAACTTTGAGGCACTGAGGCAGATTCAGTATAGTCGCTACA 2272

 821 ..................................................... 821

2273 CGGCATTCAGTCTGCAGGATCCATTTGCCCCTACCCAGCATTTAGTTTTA 2322
                                     :::||||||||||
 822 ................................SerGlnHisLeu...... 825

2323 AATCTTTTAGATGCTTGGAGAGTTCCCTTAAAAGTGTCTGCAGAGTTGTT 2372

 825 ..................................................... 825

2373 GATGAGTATAAGACTACCTCCCAAACTGAAACAGAATGAAGTAGCTAATG 2422

 825 ..................................................... 825

2423 TCCAGCCTTCTTCCCAAAGAGCCAGAATTGAAGACGTACCACCTCCCACT 2472
                                       |||||||||
 826 ........................AspGluAspValThrLysHisGln 833

2473 AAAAAGCTAACTCCAGAATTGACCCCTTTTGTGCTTTTCACTGGATTCGA 2522
     ||||||:::          |||:::   |||   :::
 834 LysLysIle.........LeuAlaArgPheAspIle.............. 842

2523 GCCTGTCCAGGTTCAACAGTATATTAAGAAGCTCTACATTCTTGGTGGAG 2572
                                                      :
 843 .............................................S 843

2573 AGGTTGCGGAGTCTGCACAGAAGTGCACACACCTCATTGCCAGCAAAGTG 2622
     ::   |||:::|||   :::::::::|||||||:::||||||::::::
 843 erGluAlaSerSerMetLysGluAlaThrHisPheIleAlaAspAsnPhe 859

2623 ACTCGCACCGTGAAGTTCCTGACGGCGATTTCTGTCGTGAAGCACATAGT 2672
     |||||||||   :::::::|||   ||||||:::      |||   :::||

 860 ThrArgThrArgAsnMetLeuGluAlaIleAlaSerGlyLysProValVa 876

2673 GACGCCAGAGTGGCTGGAAGAATGCTTCAGGTGTCAGAAGTTCATTGATG 2722
     ||||   :::||||||||||:::       :::   :::   ::::::||||
 876 lThrThrGlnTrpLeuGluSerIleAspGlnValAsnIleTyrValAspG 893

2723 AGCAGAACTACATTCTCCGAGATGCTGAGGCAGAAGTACTTTTCTCTTTC 2772
     ||:::   ||||||||||||||||:::::   |||      |||   |||
 893 luAspMetTyrIleLeuArgAspSerLysLysGluLysGluPheCysPhe 909

2773 AGCTTGGAAGAATCCTTAAAACGGGCACACGTTTCTCCACTCTTTAAGGC 2822
     ::::::      ||||||   |||||||:::      ||||||:::::::
 910 AsnMetGlyValSerLeuAlaArgAlaArgGlnProProLeuLeuGlnGl 926

2823 AAAATATTTTTACATCACACCTGGAATCTGCCCAAGTCTTTCCACTATGA 2872
     ::::      :::||||||||||:::      |||:::|||:::|||:::
 926 yArgArgValPheIleThrProAsnThrLysProAlaLeuAsnThrIleT 943

2873 AGGCAATCGTAGAGTGTGCAGGAGGAAAGGTGTTATCCAAGCAGCCATCT 2922
       ::::::|||:::::::::   |||
 943 hrThrLeuValLysAlaValHisGlyLeuPro.................. 953

2923 TTCCGGAAGCTCATGGAGCACAAGCAGAACTCGAGTTTGTCGGAA..... 2967
                  :::|||:::      :::||||||||||||||||
 954 ............ValGluArgLeuGlyArgSerSerLeuSerGluAspLy 966

2968 ............ATAATTTTAATATCCTGTGAAAATGACCTTCATTTAT 3004
                  :::::::::::::||||||||||:::|||      :::|
 966 sValProGluAsnLeuLeuValLeuSerCysGluGluAspArgAlaIleC 983

3005 GCCGAGAATATTTTGCCAGAGGCATACGATGTTCACAATGCAGAGTTCGTT 3054
     ||      ::::::   ||||||   :::|||::::::::::||||::::::
 983 ysIleProPheLeuGluArgGlyAlaGluValTyrSerSerGluLeuLeu 999

3055 CTGACTGGAGTGCTCACTCAAACGCTGGACTATGAATCATATAAG 3099
     |||:::|||::::::||||||   |||:::||||||   |||:::
1000 LeuAsnGlyIleValThrGlnArgLeuGluTyrGluArgTyrArg 1014
```

seq_name: sp_rodent:Q91VX3

```
seq_documentation_block:
ID   Q91VX3      PRELIMINARY;      PRT;  1296 AA.
AC   Q91VX3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SIMILAR TO TOPOISOMERASE (DNA) II BINDING PROTEIN
DE   (FRAGMENT).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BREAST TUMOR;
RA   Strausberg R.;
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC007170; AAH07170.1; -.
KW   Isomerase.
FT   NON_TER       1      1
SQ   SEQUENCE   1296 AA;  143663 MW;  B118FCC8B21A02D6 CRC64;


alignment_scores:
            Quality:  283.50              Length:    869
              Ratio:  0.725                Gaps:     41
 Percent Similarity:  44.994   Percent Identity: 22.325

alignment_block:
 US-09-664-641-10 x Q91VX3    ..

 Align seg 1/1  to: Q91VX3  from: 1  to: 1296
```

```
 512 GlyLysArgAlaAspGluAsnHisPheLeuValAspAsnAlaProLysGl 528

2238 ..................................................... 2238

 528 nGluGlnValLeuGluThrLysIleProAsnGlyValSerSerAsnProA 545

2239 .....................................GAGGCA 2244
                                           ::::||
 545 spLeuProAlaHisProAspAlaHisLeuGluIleHisArgLysLysAla 561

2245 CTGAGGCAGATTCAGTATAGTCGCTACACGGCATTCAGTCTG..... 2286
      :::     ::::::   :::|||:::   :::   ::::::
 562 ValThrProLeuAspMetAsnArgPheGlnSerArgAlaPheArgAlaVa 578

2287 ..................CAGGATCCATTT.................. 2298
                        |||||||||
 578 lIleSerGlnGlnArgGlyGlnAspProThrPheProProValArgGlnP 595

2299 ..........GCCCCTACCCAGCATTTA..................... 2316
                 |||:::   ||||||
 595 roLeuThrLysGluProSerLeuHisLeuAspThrProSerLysPheLeu 611

2317 ..............................GTTTTAAATCTTTTAGA 2333
                                    |||   :::   |||
 612 SerLysAspLysLeuPheLysProSerPheAspValThrAspAlaLeuAl 628

2334 TGCTTGGAGAGTTCCCTTAAAAGTGTCTGCAGAGTTGTTGATGAGTATAA 2383
       |||   :::::::|||      :::|||   :::      :::|||
 628 aAlaLeuGluThrProAsnAlaAlaSerGlnLysArgLysLeuSerSerP 645

2384 GACTACCTCCCAAACTGAAACAG.....................AATGAA 2412
       |||         :::   :::                      |||:::
 645 roLeuSerGluValIleValArgAsnLeuThrValAlaLeuAlaAsnSer 661

2413 GTAGCTAATGTCCAGCCTTCTTCCCAAAGAGCCAGAATTGAAGACGTACC 2462
          |||::::::      |||   :::|||:::||||||      |
 662 SerArgAsnThrAspSerHisSer.AlaSerProGlnLeuLysGlyAlaH 678

2463 ACCTC.....................CCACTAAAAAGCTAACTCCAGAAT 2491
     |||||                    ||||||   |||
 678 isLeuGluGluGluGluThrArgLysProLeuAspSer............ 690

2492 TGACCCCTTTTGTGCTTTTCA......CTGGATTCGAGCCTGTCCAGGTT 2535
           :::::|||   |||       |||::::::::::   |||:::::
 691 ...ValValValCysValSerLysLysLeuSerLysLysGlnSerGluLe 706

2536 CAACAGTATATTAAGAAGCTCTACATTCTTGGTGGAGAGGTTGCGGAGTC 2585
     :|||::::::      ||| |||   :::   :::|||
 706 uAsnGlyValAlaAlaSer.LeuGlyAlaGluTyrArgTrpSerPheAsp 722

2586 TGCACAGAAGTGCACACACCTCAT...TGCCAGCAAAGTGACTCGCACCG 2632
       |||   :::|||      :::      :::   :::|||||||:::||
 723 GluThrValThrHisPheIleTyrGlnGlyArgAlaAsnAspSerAsnAr 739

2633 TGAAGTTCCTGACGGCGATTTCTGTCGTGAAGCACATAGTGACGCCAGAG 2682
     |||| :::   :::|||   :::          ||||||||||:::   :::
 739 gGlu.TyrLysSerAlaLysGluArgGlyValHisIleValSerGluHis 755

2683 TGGCTGGAAGAATGCTTCAGGTGTCAGAAGTTCATTGATGAGCAGAACTA 2732
     ||||||   ||||||   :::      |||   :::   |||:::   ||
 756 TrpLeuLeuGluCysAlaGlnGluTyrLysHisLeuProGluSerLeuTy 772

2733 C 2733
     |
 772 r 772

seq_name: sp_plant:Q9ZR18

seq_documentation_block:
ID   Q9ZR18      PRELIMINARY;      PRT;   765 AA.
AC   Q9ZR18;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   HYPOTHETICAL 84.9 KDA PROTEIN.
GN   F4C21.5 OR AT4G03130.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Habermann K., de la Bastide M., Huang E.N., Gnoj L., Schutz K.,
RA   Preston R., Calma C., Martienssen R., Parnell L.D., Dedhia N.,
RA   McCombie W.R.;
RT   "Arabidopsis thaliana BAC F4C21 from chromosome IV near 17 cM.";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Spiegel L.A., Huang E.N., Nascimento L.U., de la Bastide M., Vil D.M.,
RA   Preston R.R., Matero A., Shah R., O'Shaughnessy A., Rodriguez M.,
RA   Chekher M., Schutz K., See L.H., Swaby I., Habermann K., Dedhia N.N.,
RA   Newes H.W., Lemcke K., Mayer K.F.X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC005275; AAD14441.1; -.
DR   EMBL; AL161496; CAB77798.1; -.
DR   InterPro; IPR001357; BRCT.
DR   Pfam; PF00533; BRCT; 2.
DR   SMART; SM00292; BRCT; 1.
DR   PROSITE; PS50172; BRCT; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   765 AA;  84874 MW;  A49BB5DADE6948E0 CRC64;


alignment_scores:
           Quality:  265.50              Length:    910
             Ratio:  0.676                Gaps:     34
Percent Similarity:  43.187   Percent Identity: 19.890

alignment_block:
US-09-664-641-10 x Q9ZR18   ..

Align seq 1/1  to: Q9ZR18  from: 1  to: 765

 766 GTTCTGGATTGC......GTATCAGAGAAAACCAAAAAGGACGAAGCATT 809
     :::|||   |||      :::::::::              |||||||||::
  26 IleLeuGlyCysGlyGluIleGluAspThrGluAlaIleAspGluAlaLe 42

 810 TTAT.....................CATCCTCGTCTGATTATTTATGAAG 838
     :                          |||      ::::::|||::::
  42 uValLeuAsnSerSerLeuLeuGluThrProPheGlnValLeuTyrAspA 59

 839 AGGAACAA.....................GAGGAAGAGGAAGAGGAG 864
     ::   |||                   :::||||||:::   :::
  59 spThrGluLeuValAspAsnGlyAsnCysGlyAspGluGluSerLeuAsn 75

 865 GAGGAAGTAGAAAATGAGGAACAA.............GATTCTCAGAATGA 902
     :::|||      :::   :::|||              ||||||::::::||
  76 GlnGluCysTyrGlyLeuAspGlnValValValAspAspSerAspAspGl 92

 903 GGGTAGTACAGATGAGAAGTCAAGCCCTGCCAGCTCTCAAGAAGGGTCTC 952
     |   ::::::          ::::::|||             |||   ||||
  92 uAspGlySerValAlaAlaGlyGluProMetCysLeuProGluLeuSerP 109

 953 CT...TCAGGTGAC..................................... 963
     ||   :::||||||
 109 roLysAlaGlyAspIleLeuLeuGluSerAspGlySerAsnAspHisGlu 125

 964 ...CAGCAGTTTTCACCTAAATCCAAC..................... 987
        |||      :::   :::||||||
 126 CysGlnThrGlyLysGlnGluSerAsnCysAspThrValThrGlyPheGl 142
```

```
 988 ..............ACTGAAAAATCTAAAGGGGAA...TTAATGTTTG 1018
                     |||  :::|||  |||     |||  |||
 142 nGlySerLeuArgIleThrAlaSerSerHisGlyProGlyLeuAspPheL 159

1019 ATGATTCTTCAGATTCATCACCGGAAAAACAG..................  1050
       ||||||::::::   :::  :::  |||
 159 euAspSerGlnGluProGlyGluSerThrGlnAlaAspAlaLeuGlyPhe 175

1051 .................GAGAGAAATTTAAACTGGACCCCGGCCGAAGT 1082
                      :::::::::||||||   :::|||:::::::
 176 ValAspHisPheLeuMetAspLysAspLeuAsnLeuSerProValAspLe 192

1083 CCCACAGTTAGCTGCAGCAAAACGCAGGCTGCCTCAGGGAAAGGAGACCTG 1132
     :|||     ::::::  ::::::     |||  ::  |
 192 uProValAsnSerCysArgArgLysSerSerProValSerGlyAlaLysG 209

1133 GGTTGATTAACTTGTGTGCCAATGTCCCACCCGTCCCAGGTAACATTTTG 1182
     ||      :::|||:::  ::::::
 209 lyCysGlnSerLeuAlaLysArgIleMetSer..................  219

1183 CCCCCTGAGGTCCCGGGTAATTTAATGGCTGCTGGACAAAACCTCCAAAG 1232
                                                 ::::|
 220 ..............................................ArgSe 221

1233 TTCTGAAAGATCAGAAATGATAGCTACCTGG.................  1263
     |     |||:::     :::     |||
 221 rProThrArgLysMetSerValPheAspTrpSerSerAspGlnCysAspV 238

1264 .................AGTCCAGCTGTACGG.........ACACTGAGG 1287
                     ||||||:::::|||         |||
 238 alSerAspProLysThrSerProValThrArgAlaSerIleThrCysSer 254

1288 AATATTACTAATAATGCT............GACATTCAGCAGATGAACCC 1325
     ::::||  ::::::              |||:::        :::||
 255 LysIleArgGluAspGlnValAlaGlyAspAspPheGlyValLysLysAr 271

1326 GCCATCAAATGTAGCACATATCTTACAGACTCTTTCAGCACCTACGAAAA 1375
     |     :::  :::  |||  ::::::  |||  ||||||:::
 271 gSerArgAspProCysAspIleArgLysValSerSerHisProThrGlnA 288

1376 ATTTAGAACAGCAGGTGAATCACAGCCAGCAGGGACATACAAATGCCAAT 1425
     :::::  :::                          :::::::::|||:::
 288 rgValMetLys.....................AsnSerSerAlaLys 296

1426 GCAGTGCTGTTTAGCCAAGTGAAAGTGACTCCAGAGACACACATGCTACA 1475
       :::  :::::::|||::::::         |||     |||:::
 297 HisLeuLysMetGluGlnAlaSerGlyLeuSerGluGlyIleMetPheIl 313

1476 GCAGCAGCAGCAGGCCCAGCAGCAGCAGCAGCACCCGGTTTTACACC 1525
      ::|||:::::|||||  :::::::::     |||       |
 313 eSerGlnLysAspAlaGlnLeuGluAspLysAla...ProLysGluProL 329

1526 TTCAGCCCCAGCAGATAATG.....................CAGCTCCAG 1554
     ||:::|||::::::  :::                     |||:::
 329 euGluProGluGluAspPheIleAspIleGlyIleAsnThrGlnIleAla 345

1555 CAGCAGCAGCAGCAGCAGATCTCTCAGCAACCTTACCCC...........  1593
       :::  ::::::  :::       |||
 346 AlaGluAlaMetSerAlaLeuValTyrAlaProCysThrThrGluGluAl 362

1594 ....CAGCAGCCGCCGCATCCATTTTCACAGCAGCAGCAGCAGCAACAGC 1639
       ::::::  |||  |||  |||  ::::::::::::::::|||
 362 aCysGluSerAspProIleProGlySerValSerGluMetArgAspGlnV 379

1640 AGCCACCACCATCGCCTCAGCAGCATCAGCTTTTTGGACATGATCCAGCA 1689
               |||  ::::::::::::  :::      :::|||
 379 alSerAsnLeuSerArgArgAsnAsnAspThrIleGluGlyGluPro... 394

1690 GTGGAGATTCCAGAAGAAGGCTTCTTATTGGGATGTGTGTTTGCAATTGC 1739
                  |||        |||:::  :::|||:::  ||
 395 ..............GluArgAspLysLeuAsnGlyLeuPheSerAlaAl 406
```

```
1740 GGATTATCCAGAGCAGATGTCTGATAAGCAACTGCTGGCCACCTGGAAAA 1789
     |  :::  |||:::  |||  |||:::         |||  ::::
 406 aProHisArgGluArgAsnSerLysLysLysArgLysPheThrAsnGluG 423

1790 GGATAATCCAGGCACATGGCGGCACTGTTGACCCCACCTTCACGAGTCGA 1839
     ::      :::|||:::                  :::  ::::::
 423 luArgThrArgAlaAsn.................ValSerValLysThr 433

1840 TGCACGCACCTTCTCTGTGAGAGTCAAGTCAGCAGCGCGTATGCACAGGC 1889
     |||        |||||||||  :::           :::   :::||
 434 CysLeuLeuAsnLeuCysGluTrpArg.....HisProArgAl 446

1890 AATAAGAGAAAGAAAGAGATGTGTTACTGCACACTGGTTAAACACAGTCT 1939
     |  |||:::|||                              |
 446 aLysArgSerArg.................................L 451

1940 TAAAAAAAAAAAAAATGGTACCGCCGCACCGAGCCCTTCACTTCCCAGTG 1989
     ||   ::::::     |||||||||:::|||:::
 451 euMetGlnArgHisHisValProProArgArgSerTrp..... 463

1990 GCCTTCCCACCAGGAGGAAAGCCATGTTCACAGCATATTATTTCTGTGAC 2039

 463 ..................................................  463

2040 TGGATTTGTTGATAGTGACAGAGATGACCTAAAATTAATGGCTTATTTGG 2089

 463 ..................................................  463

2090 CAGGTGCCAAATATACGGGTTATCTATGCCGCAGCAACACAGTCCTCATC 2139
       |||||||:::   :::          ::::::||||||
 464 ..GlyAlaSerLeuValLysAspArgSerGluThrAsnThr..... 476

2140 TGTAAAGAACCAACTGGTTTAAAGTATGAAAAAGCCAAAGAGTGGAGGAT 2189

 476 ..................................................  476

2190 ACCCTGTGTCAACGCCCAGTGGCTTGGCGACATTCTTCTGGGAAACTTTG 2239
                          |||::::::   |||  |||:::::::
 477 .....................LeuSerSerSerLeuArgGlySerLeuS 486

2240 AGGCACTGAGGCAGATTCAGTATAGTCGCTACACGGCATTCAGTCTGCAG 2289
     :::::  ||||||
 486 erGlyThrArgGlnAla.............................. 491

2290 GATCCATTTGCCCCTACCCAGCATTTAGTTTTAAATCTTTTAGATGCTTG 2339
               :::      |||      |||::::::::|||      |||::
 492 .........SerSerCysGlnSerGlyValIleAspLeuAsnValAlaTy 505

2340 GAGAGTT...CCCTTAAAAGTG.............................  2358
     :::::::  |||  |||:::
 505 rHisAlaSerProArgLysIleTyrAspGlySerHisGluSerProCysA 522

2359 ..........................TCTGCAGAGTTGTTGATGAGTATA 2382
                                :::  ||||||     |||:::
 522 snLysAspPheProArgLeuPheLeuGlnLysGluLeuThrThrSerLeu 538

2383 AGACTACCTCCCAAACTG.........................AAACA 2405
           |||  |||:::                          |||::
 539 GlyGlyProGlyLysIleGlyAspPheValTrpLysAspLeuArgLysAr 555

2406 GAATGAAGTAGCTAATGTCCAGCCTTCTTCCCAAAGAGCCAGAATTGAAG 2455
     ::::::::::|||:::|||:::
 555 gArgAsnLeuAlaHisValArg.............................. 562

2456 ACGTACCACCTCCCACTAAAAAGCTAACTCCAGAATTGACCCCTTTTGTG 2505
                                                  |||
 563 ..............................................Val 563

2506 CTTTTCACT...GGATTCGAGCCTGTCCAGGTTCAACAGTATATTAAGAA 2552
     ||||||:::   ::::::::        |||:::|||     |||
 564 LeuPheSerGlnAsnLeuAspAspGluThrValLysGlnGlnLysLysIl 580

2553 GCTCTACATTCTTGGTGGAGAGGTTGCGGAGTCTGCACAGAAGTGCACAC 2602
```

```
          :::      ||||||   :::   |||:::|||:::          ||||
 580 cMetValArgLeuGlyIleSerProAlaSerSerSerAlaAspSerThrH 597

2603 ACCTCATTGCCAGCAAAGTGACTCGCACCGTGAAGTTCCTGACGGCGATT 2652
     ||:::||||||::::::   :::||||||   ::::::|||   ||||||
 597 isPheIleAlaAspArgPheAlaArgThrArgAsnMetLeuGluAlaIle 613

2653 TCTGTCGTCAAGCACATAGTGACGCCACAGTGGCTGGAAGAATGCTTCAG 2702
     ::::::   |||   :::||||||||||   |||||||||:::|||   ::
 614 AlaLeuGlyLysPheValValThrProIleTrpLeuGluSerCysAlaGl 630

2703 GTGTCAGAAGTTCATTGATGAGCAGAACTACATTCTCCGAGATGCTGAGG 2752
     :   :::   :::||||||||||:::::::|||||||||||||||   :::
 630 nThrArgCysLeuIleAspGluLysSerTyrIleLeuArgAspIleLysL 647

2753 CAGAAGTACTTTTCTCTTTCAGCTTGGAAGAATCCTTAAAACGGGCACAC 2802
        |||         :::||||   |||      ||||||   ||||||
 647 ysGlu...LysAspGlyPheCysLeuLeuThrSerLeuAlaArgAlaLys 662

2803 GTTTCTCCACTCTTTAAGGCAAAATATTTTTACATCACACCTGGAATCTG 2852
           ||||||:::|||:::            ||||||||||:::|||
 663 GlnHisProLeuLeuLysGlyPheLysValCysIleThrProSerIleLy 679

2853 CCCAAGTCTTTCCACTATGAAGGCAATCGTAGAGTGTGCAGGAGGAAAGG 2902
      ||||||   :::   :::      :::|||:::   :::   |||:::|
 679 sProSerArgGlyMetIleThrAspLeuValLysMetThrGlnGlyGlnV 696

2903 TGTTATCCAAGCAGCCATCTTTCCGGAAGCTCATGGAGCACAAGCAGAAC 2952
     ||:::::::   :::           :::::::::      :::::::::
 696 alValGluAlaSer.....GluIleIleAlaAlaGluAspArg 708

2953 TCGAGTTTGTCGGAAATAATTTTAATATCCTGTGAAAATGACCTTCATTT 3002
     :::      ||||||||||||||||||||||:::::::|||      ::
 709 AsnPheProGluAspValLeuIleLeuSerCysLysGluAspArgAspPh 725

3003 ATGCCGAGAATATTTTGCCAGAGGCATA..................... 3030
     :|||      :::      :::|||
 725 eCysLeuProPheValAsnGlnGlyThrValSerSerGlnPheGlnArgL 742

3031 ..........GATGTTCACAATGCAGAGTTCGTTCTGACTGGAGTGCTC 3069
                :::::::|||      :::|||:::   :::   :::|||
 742 ysGlyThrLysGluIleHis......LysPheThrAlaAlaCysLeuLeu 756

3070 ACTCAAACGCTGGACTATGAATCATATAAG 3099
        :::|||   |||:::      ||||||
 757 LeuSerThrThrAspTrp......TyrLys 764
```

seq_name: sp_human:Q9UEB9

seq_documentation_block:

```
ID   Q9UEB9      PRELIMINARY;      PRT;  1435 AA.
AC   Q9UEB9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   DNA TOPOISOMERASE II BINDING PROTEIN.
GN   TOPBP1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Yamane K., Kawabata M., Tsuruo T.;
RT   "A DNA topoisomerase II binding protein with eight repeating regions
RT   similar to DNA repair enzymes and to a cell cycle regulator.";
RL   Eur. J. Biochem. 250:794-799(1998).
DR   EMBL; AB019397; BAA34202.1; -.
DR   InterPro; IPR001357; BRCT.
DR   Pfam; PF00533; BRCT; 7.
DR   SMART; SM00292; BRCT; 7.
DR   PROSITE; PS50172; BRCT; 7.
KW   Isomerase.
```

```
SQ   SEQUENCE   1435 AA;  160550 MW;  39FF4399A076AF9B CRC64;
```

alignment_scores:

```
           Quality:  263.50              Length:   1031
             Ratio:  0.620                 Gaps:     43
Percent Similarity:  41.222   Percent Identity: 19.011
```

alignment_block:

US-09-664-641-10 x Q9UEB9 ..

Align seg 1/1 to: Q9UEB9 from: 1 to: 1435

```
 234 AATGCACTAGCCTCACACATAATCTCAGAGGATGGGGACAATCCAGAGGT 283
     |||:::   :::::::|||:::|||:::                 |||||
  53 AsnValSerValThrHisLeuIleAla.............GlyGluVa 64

 284 GGGAGAAGCTCGGGAAGTC...........TTTGACTTACCTGTTGTAA 321
     ||||:::   :::   :::           :::     |||::::::
  64 lGlySerLysLysTyrLeuValAlaAlaAsnLeuLysLysProIleLeuL 81

 322 AGCCTTCTTGGGTGATTCTG.....................TCCGTT 347
       ||||||||||:::                          ::::::
  81 euProSerTrpIleLysThrLeuTrpGluLysSerGlnGluLysLysIle 97

 348 CAGTGTGGAACTCTTCTGCCAGTAAATGGTTTTTCTCCAGAATCATGTCA 397
         |||   :::   :::::   |||         :::|||
  98 ThrArgTyrThrAspIleAsnMetGluAspPhe.....LysCysPr 111

 398 GATTTTTTTTGGAATCACTGCCTGCCTTTCTCAGGGTGTTGATACAAGCT 447
     |||||||:::|||         |||::::::
 111 oIlePheLeuGlyCysIleIleCysValThr..................  121

 448 GGAGCTCTTTGTTGGAGTCTTCCAGAGCTCTCCCAAGGAGAGGTAGGGAA 497

 121 .................................................  121

 498 GGGAGCTTGTCCAGCAGAAGTTGGGAAGCACAGAGATCATCTGCCTTCTT 547

 121 .................................................  121

 548 CTGACCCGGTATTGATGCAGGCTGAGGCCTCTGTTGTAATGTGCTGGGTG 597
                                 ||||||||   :::
 122 ........................GlyLeuCys.GlyLeu.........  126

 598 TCATCTGAAGACAGAAGTGCCCTGTGGGCTTTGGTTACGTTCTATGGGGG 647
               ||||||:::   :::      |||::::::   :::|||||
 127 .........AspArgLysGluValGlnGlnLeuThrValLysHisGlyGl 140

 648 AGATTGC......CAGCTAACCCTCAATAAGAAATGCACGCATTTGATTG 691
     |:::         ||||||   :::|||   :::|||||||||||||||||
 140 yGlnTyrMetGlyGlnLeuLysMetAsn...GluCysThrHisLeuIleV 156

 692 TTCCAGAGCCAAAGGGGGGGAAATACGAATGTGCTTTAAAGCGAGCAAGT 741
     ||   ||||||||||||   ||||||||||||||||   ||||||   :::
 156 alGlnGluProLysGlyGlnLysTyrGluCysAla...LysArgTrpAsn 171

 742 ATTAAAATTGTGACTCCTGACTGGGTTCTGGATTGCGTATCAGAGAAAAC 791
     :::      ||||||   :::|||   :::|||   ::::::::
 172 ValHisCysValThrThrGlnTrpPhePheAspSerIleGluLysGlyPh 188

 792 CAAAAAGGACCAAGCATTTTATCATCCTCGTCTGATTATTTATGAAGAGG 841
         :::||||||:::::|||                        :
 188 eCysGlnAspGluSerIleTyr.........................L 196

 842 AAGAAGAGGAAGAGGAAGAGGAGGAGGAAGTAGAAAATGAGGAACAAGAT 891
     ::   |||   :::   |||   :::   :::   |||:::::
 196 ysThrGluProArgProGluAlaLysThrMetProAsnSerSerThrPro 212

 892 TCTCAGAATGAGGGTAGTACAGATGAGAAGTCAAGCCCTGCCAGCTCTCA 941
     :::::::::   ::::::   |||::::::::::
 213 ThrSerGlnIleAsnThrIleAspSerArgThr.................  223
```

```
 942 AGAAGGGTCTCCTTCAGGTGACCAGCAGTTTTCACCTAAATCCAACACTG 991
                                 :::|||      ||||||   :
 224 .........................LeuSerAspValSerAsnIleS 231

 992 AAAAATCTAAAGGGGAATTAATGTTTGATGATTCTTCAGATTCATCACCG 1041
     :::::   :::::::::   :::   ::::::      :::|||
 231 erAsnIleAsnAlaSerCysValSerGluSerIleCysAsnSerLeuAsn 247

1042 GAAAAACAGGACAGAAATTTAAACTGGACCCCGGCCGAAGTCCCACAGTT 1091
     :::|||   |||   :::|||:::                ::::::  :::||
 248 SerLysLeuGluProThrLeuGlu.............AsnLeuGluAsnLe 260

1092 AGCTGCAGCAAAACGCAGGCTGCCTCAGGGAAAGGAGCCTGGGTTG...A 1138
     |   ::::::      :::   |||:::            |||      |
 260 uAspValSerAlaPheGlnAlaProGluAspLeuLeuAspGlyCysArgI 277

1139 TTAACTTGTGTGCCAATGTCCCACCCGTCCCAGGTAACATTTTGCCCCCT 1188
     ||   ||||||:::
 277 leTyrLeuCysGly..................................... 281

1189 GAGGTCCGGGGTAATTTAATGGCTGCTGGACAAAACCTCCAAAGTTCTGA 1238
                              :::|||:::::::|||:::::
 282 ......................PheSerGlyArgLysLeuAspLys..... 289

1239 AAGATCAGAAATGATAGCTACCTGGAGTCCAGCTGTACGGACACTGAGGA 1288
                                       :::|||   |||   |
 290 ................................LeuArgArgLeuIleA 295

1289 ATATTACTAATAATGCTGACATTCAGCAGATGAACCGGCCATCAAATGTA 1338
     ||      :::::::::   ::::::|||:::|||      :::::::|||
 295 snSerGlyGlyGlyValArgPheAsnGlnLeuAsn......GluAspVal 309

1339 GCACATATCTTACAGACTCTTTCAGCACCTACGAAAAATTTAGAACAGCA 1388
     :::|||::::::      ::::::              :::|||:::|||
 310 ThrHisValIle......ValGlyAspTyrAspAspGluLeuLysGlnPh 324

1389 GGTGAATCACAGC...CAGCAGGGACATACAAATGCCAATGCAGTGCTGT 1435
        |||   |||   ::::::   |||:::   :::         |||
 324 eTrpAsnLysSerAlaHisArgProHisValValGlyAlaLysTrpLeu. 340

1436 TTAGCCAAGTGAAAGTGACTCCAGAGACACACATGCTACAGCAGCAGCAG 1485
           ::::::         :::   :::||||||:::::::::
 341 ........LeuGluCysPheSerLysGlyTyrMetLeuSerGluGluPro 354

1486 ......CAGGCCCAGCAGCAGCAGCAGCAGCACCCGGTTTTACACCTTCA 1529
           :::|||:::   |||      :::  ||||||   |||   ||
 355 TyrIleHisAlaAsnTyrGlnProValGluIleProValSerHis...Gl 370

1530 GCCCCAGCAGATAATGCAGCTCCAGCAGCAGCAGCAGCAGCAGATCTCTC 1579
     ||||::::::        |||   :::::::::::::::::::::|||:
 370 nProGluSerLysAlaAlaLeuLeuLysLysLysAsnSerSerPheSerL 387

1580 AGCAACCTTACCCCCAGCAGCCGCCGCAT..................... 1608
     :::::   :::      :::      |||
 387 ysLysAspPheAlaProSerGluLysHisGluGlnAlaAspGluAspLeu 403

1608 ................................................... 1608

 404 LeuSerGlnTyrGluAsnGlySerSerThrValValGluAlaLysThrSe 420

1609 .........CCATTTTCACAGCAGCAGCAGCAGCAA.............. 1635
             ||||||:::::::::   :::   :::
 420 rGluAlaArgProPheAsnAspSerThrHisAlaGluProLeuAsnAspS 437

1636 .................CAGCAGCCACCACCATCGCCTCAGCAGCATCAG 1668
                     |||:::         |||      :::|||
 437 erThrHisIleSerLeuGlnGluGluAsnGlnSerSerValSerHisCys 453

1669 CTTTTTGGACATGATCCAGCAGTGGAG...ATTCCAGAAGAAGGCTTCTT 1715
     :::            |||   |||:::   |||   ||||||||||:::::
 454 Val............ProAspValSerThrIleThrGluGluGlyLeuPh 466

1716 ATTGGGATGTGTGTTTGCAATTGCGGATTATCCAGAGCAGATGTCTGATA 1765

     :           |||   :::      :::   ::::::   :::
 466 eSerGlnLysSerPheLeuValLeuGlyPheSerAsnGluAsnGlu.... 481

1766 AGCAACTGCTGGCCACCTGGAAAAGGATAATCCAGGCACATGGCGGCACT 1815
                    ::::::      :::|||||||:::  ::::::|||
 482 ..........SerAsnIleAlaAsnIleIleLysGluAsnAlaGlyLys 494

1816 GTTGACCCCACCTTCACGAGTCGATGCACGCAC.................. 1848
     :::      ::::::   ||||||   :::
 495 Ile...MetSerLeuLeuSerArgThrValAlaAspTyrAlaValValPr 510

1849 ....CTTCTCTGTGAGAGTCAAGTCAGCAGCGCGTATGCACAGGCAATAA 1894
         |||   ||||||   :::::::::   :::
 510 oLeuLeuGlyCysGluValGluAlaThrValGly................. 521

1895 GAGAAAGAAAGAGATGTGTTACTGCACACTGGTTAAACACAGTCTTAAAA 1944
                ::::   ||||||      ||||||   |||   :::
 522 ..........GluValValThrAsnThrTrpLeuValThrCysIleAsp 534

1945 AAAAAAAAAATGGTA...CCGCCGCACCGAGCCCTTCACTTCCCAGTGGC 1991
        :::   :::      |||      :::   |||      ||||||
 535 TyrGlnThrLeuPheAspProLysSerAsnProLeuPheThrProValPr 551

1992 CTTCCCACCAGGAGGAAAGCCATGTTCACAGCATATTATTTCTGTGACTG 2041
                |||      |||   ::::::   :::|||||||   :::
 551 oValMetThrGlyMetThrProLeuGluAspCysValIleSerPheSerG 568

2042 GATTTGTTGATAGTGACAGAGATGACCTAAAATTAATGGCTTATTTGGCA 2091
        :::   ::::::::::::::|||   ::::::|||   |||
 568 lnCysAlaGlyAlaGluLysGluSerLeuThrPheLeuAlaAsnLeuLeu 584

2092 GGTGCCAAATATACGGGTTATCTATGCCGC..................... 2121
     ||||||:::      |||:::   |||
 585 GlyAlaSerVa'GlnGluTyrPheValArgLysSerAsnAlaLysLysGl 601

2122 .......AGCAACACAGTCCTCATCTGTAAAGAACCAACTGGTTTAAAGT 2164
              ::::::|||   ||||||   ||||||      |||   ||||
 601 yMetPheAlaSerThrHisLeuIleLeuLysGluArgGlyGlySerLysT 618

2165 ATGAAAAAGCCAAAGAGTGGAGGATACCCTGTGTCAACGCCCAGTGGCTT 2214
     |||||   ||||||:::|||:::::::|||:::|||:::      ||||||
 618 yrGluAlaAlaLysLysTrpAsnLeuProAlaValThrIleAlaTrpLeu 634

2215 ..................................GGCGACATTCTTCT 2228
                                        :::   :::||||::
 635 LeuGluThrAlaArgThrGlyLysArgAlaAspGluSerHisPheLeuIl 651

2229 GGGAAAC.............................................. 2235
     :   |||
 651 eGluAsnSerThrLysGluGluArgSerLeuGluThrGluIleThrAsnG 668

2235 ................................................... 2235

 668 lyIleAsnLeuAsnSerAspThrAlaGluHisProGlyThrArgLeuGln 684

2236 .....................................................TT 2237

 685 ThrHisArgLysThrValValThrProLeuAspMetAsnArgPheGlnSe 701

2238 TGAGGCACTGAGGCAGATT......CAGTATAGTCGCTACACGGCATTCA 2281
     :::|||:::|||   :::      |||::::::|||   :::|||   |
 701 rLysAlaPheArgAlaValValSerGlnHisAlaArgGlnValAlaAlaS 718

2282 GT.................................................. 2283
     ||
 718 erProAlaValGlyGlnProLeuGlnLysGluProSerLeuHisLeuAsp 734

2283 ................................................... 2283

 735 ThrProSerLysPheLeuSerLysAspLysLeuPheLysProSerPheAs 751

2284 .....................CTGCAGGATCCATTTGCCCCTACCCAGC 2311
                          |||:::   |||      |||:::||||:
```

```
 751 pValLysAspAlaLeuAlaAlaLeuGluThrProGlyArgProSerGlnG 768

2312 AT..............................TTAGTTTTAAATCTT 2328
     ::                             :::|||   ||||||
 768 lnLysArgLysProSerThrProLeuSerGluValIleValLysAsnLeu 784

2329 TTAGATGCTTGGAGAGTTCCCTTAAAAGTGTCTGCAGAGTTGTTGATGAG 2378
              ::::::   |||      |||::::::      :::::
 785 ............GlnLeuAlaLeuAlaAsnSerSerArgAsnAlaValAl 797

2379 TATAAGACTACCTCCCAAACTGAAACAGAATGAAGTAGCTAATGTCCAGC 2428
     ::::         |||:::|||||::
 797 aleuSerAlaSerProGlnLeuLysGlu..................... 806

2429 CTTCTTCCCAAAGAGCCAGAATTGAAGACGTACCACCTCCCACTAAAAAG 2478
              :::::::::   |||:::::::||   |||      |||
 807 ........AlaGlnSerGluLysGluGluAlaProLysProLeuHisLys 820

2479 CTAACTCCAGAATTGACCCCTTTTGTGCTTTTCACTGGATTCGAGCCTGT 2528
     ::::::      ::::::
 821 ValValValCysValSer............................... 826

2529 CCAGGTTCAACAGTATATTAAGAAGCTCTACATTCTTGGTGGAGAGGTTG 2578
            ::::::      ||||||          |||:::|||   :::|
 827 .......LysLysLeuSerLysLysGlnSerGluLeuAsnGlyIleAlaA 841

2579 CGGAGTCTGCA.......................CAGAAGTGCACACAC 2604
     ||:::   :::                      :::      ||||||
 841 laSerLeuGlyAlaAspTyrArgTrpSerPheAspGluThrValThrHis 857

2605 CTCATT............GCCAGCAAAGTGACTCGCACCGTGAAGTTCCT 2642
     :::|||                :::   ::::::|||      |||   ::
 858 PheIleTyrGlnGlyArgProAsnAspThrAsnArgGluTyrLysSerVa 874

2643 GACGGCGATTTCTGTCGTGAAGCACATACTCAGCCAGAGTGGCTGGAAG 2692
     :         :::|||      ||||||||:::   :::||||||   :
 874 lLysGluArgGlyVal......HisIleValSerGluHisTrpLeuLeuA 889

2693 AATGCTTCAGGTGTCAGAAGTTCATTGATGAGCAGAACTAC 2733
     ::|||   :::      |||   :::   |||:::   |||
 889 spCysAlaGlnGluCysLysHisLeuProGluSerLeuTyr 902
```

seq_name: sp_human:Q92547

seq_documentation_block:

ID Q92547 PRELIMINARY; PRT; 1550 AA.
AC Q92547;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE KIAA0259 PROTEIN (FRAGMENT).
GN KIAA0259.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BONE MARROW;
RX MEDLINE=97191544; PubMed=9039502;
RA Nagase T., Seki N., Ishikawa K., Ohira M., Kawarabayasi Y., Ohara O.,
RA Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. VI.
RT the coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT analysis of cDNA clones from cell line KG-1 and brain.";
RL DNA Res. 3:321-329(1996).
DR EMBL; D87448; BAA13389.1; -.
DR InterPro; IPR001357; BRCT.
DR Pfam; PF00533; BRCT; 8.
DR SMART; SM00292; BRCT; 7.
DR PROSITE; PS50172; BRCT; 7.
FT NON_TER 1 1
SQ SEQUENCE 1550 AA; 173388 MW; 86F6B66AFB6B292D CRC64;

alignment_scores:

Quality: 263.50 Length: 1031
Ratio: 0.620 Gaps: 43
Percent Similarity: 41.222 Percent Identity: 19.011

alignment_block:
US-09-664-641-10 x Q92547 ..

Align seg 1/1 to: Q92547 from: 1 to: 1550

```
 234 AATGCACTAGCCTCACACATAATCTCAGAGGATGGGGACAATCCAGAGGT 283
     |||:::   ::::::|||:::|||:::                    |||||
 168 AsnValSerValThrHisLeuIleAla..............GlyGluVa 179

 284 GGGAGAAGCTCGGGAAGTC............TTTGACTTACCTGTTGTAA 321
     ||||:::   :::   :::            :::     |||::::::
 179 lGlySerLysLysTyrLeuValAlaAlaAsnLeuLysLysProIleLeuL 196

 322 AGCCTTCTTGGGTGATTCTG.......................TCCGTT 347
       |||||||||:::                             ::::::
 196 euProSerTrpIleLysThrLeuTrpGluLysSerGlnGluLysLysIle 212

 348 CAGTGTGGAACTCTTCTGCCAGTAAATGGTTTTTCTCCAGAATCATGTCA 397
               |||   :::   ::::::   |||         :::|||
 213 ThrArgTyrThrAspIleAsnMetGluAspPhe.........LysCysPr 226

 398 GATTTTTTTTGGAATCACTGCCTGCCTTTCTCAGGGTGTTGATACAAGCT 447
     ||||||:::|||         |||::::::
 226 oIlePheLeuGlyCysIleIleCysValThr.................... 236

 448 GGAGCTCTTTGTTGGAGTCTTCCAGAGCTCTCCCAAGGAGAGGTAGGGAA 497

 236 ..................................................... 236

 498 GGGAGCTTGTCCAGCAGAAGTTGGGAAGCACAGAGATCATCTGCCTTCTT 547

 236 ..................................................... 236

 548 CTGACCCGGTATTGATGCAGGCTGAGGCCTCTGTTGTAATGTGCTGGGTG 597
                                 |||||||||    :::
 237 .......................GlyLeuCys.GlyLeu......... 241

 598 TCATCTGAAGACAGAAGTGCCCTGTGGGCTTTGGTTACGTTCTATGGGGG 647
           ||||||:::   :::     |||:::::   :::|||||
 242 .........AspArgLysGluValGlnGlnLeuThrValLysHisGlyGl 255

 648 AGATTGC......CAGCTAACCCTCAATAAGAAATGCACGCATTTGATTG 691
     |:::         ||||||   :::|||   :::||||||||||||||||
 255 yGlnTyrMetGlyGlnLeuLysMetAsn...GluCysThrHisLeuIleV 271

 692 TTCCAGAGCCAAAGGGGGGGAAATACGAATGTGCTTTAAAGCGAGCAAGT 741
     ||   ||||||||||||   ||||||||||||||   ||||||   :::
 271 alGlnGluProLysGlyGlnLysTyrGluCysAla...LysArgTrpAsn 286

 742 ATTAAAATTGTGACTCCTGACTGGGTTCTGGATTGCGTATCAGAGAAAAC 791
     :::      ||||||   :::|||   :::|||   :::::::::
 287 ValHisCysValThrThrGlnTrpPhePheAspSerIleGluLysGlyPh 303

 792 CAAAAAGGACGAAGCATTTTATCATCCTCGTCTGATTATTTATGAAGAGG 841
       :::|||||||:::::|||                            :
 303 eCysGlnAspGluSerIleTyr.........................L 311

 842 AAGAAGAGGAAGAGGAAGAGGAGGAGGAAGTAGAAAATGAGGAACAAGAT 891
     ::   |||   :::   |||   :::   :::   |||:::::::
 311 ysThrGluProArgProGluAlaLysThrMetProAsnSerSerThrPro 327

 892 TCTCAGAATGAGGGTAGTACAGATGAGAAGTCAAGCCCTGCCAGCTCTCA 941
     :::::::::   ::::::   |||:::::::::
 328 ThrSerGlnIleAsnThrIleAspSerArgThr.................. 338

 942 AGAAGGGTCTCCTTCAGGTGACCAGCAGTTTTCACCTAAATCCAACACTG 991
```

```
 339 ..........................LeuSerAspValSerAsnIleS 346

 992 AAAAATCTAAAGGGGAATTAATGTTTGATGATTCTTCAGATTCATCACCG 1041
     :::::   :::::::::   :::   ::::::       :::|||
 346 erAsnIleAsnAlaSerCysValSerGluSerIleCysAsnSerLeuAsn 362

1042 GAAAAACAGGAGAGAAATTTAAACTGGACCCCGGCCGAAGTCCCACAGTT 1091
     :::|||   |||   :::|||:::            ::::::   :::||
 363 SerLysLeuGluProThrLeuGlu.............AsnLeuGluAsnLe 375

1092 AGCTGCAGCAAAACCCAGGCTGCCTCAGGCAAAGGAGCCTGGGTTG...A 1138
     |   ::::::      :::   |||:::              |||     |
 375 uAspValSerAlaPheGlnAlaProGluAspLeuLeuAspGlyCysArgI 392

1139 TTAACTTGTGTGCCAATGTCCCACCCGTCCCAGGTAACATTTTGCCCCCT 1188
     ||   ||||||:::
 392 leTyrLeuCysGly..................................... 396

1189 GAGGTCCGGGGTAATTTAATGGCTGCTGGACAAAACCTCCAAAGTTCTGA 1238
                          :::|||:::::::|||:::::
 397 ....................PheSerGlyArgLysLeuAspLys..... 404

1239 AAGATCAGAAATGATAGCTACCTGGAGTCCAGCTGTACGGACACTGAGGA 1288
                                    :::|||   |||   |
 405 ................................LeuArgArgLeuIleA 410

1289 ATATTACTAATAATGCTGACATTCAGCAGATGAACCGGCCATCAAATGTA 1338
     ||      :::::::::   ::::::|||:::|||      ::::::|||
 410 snSerGlyGlyGlyValArgPheAsnGlnLeuAsn......GluAspVal 424

1339 GCACATATCTTACAGACTCTTTCAGCACCTACGAAAAATTTAGAACAGCA 1388
     :::|||::::::      ::::::             :::|||:::|||
 425 ThrHisValIle......ValGlyAspTyrAspAspGluLeuLysGlnPh 439

1389 GGTGAATCACAGC...CAGCAGGGACATACAAATGCCAATGCAGTGCTGT 1435
         |||   |||   ::::::   |||:::   :::         |||
 439 eTrpAsnLysSerAlaHisArgProHisValValGlyAlaLysTrpLeu. 455

1436 TTAGCCAAGTGAAAGTGACTCCAGAGACACACATGCTACAGCAGCAGCAG 1485
            ::::::          :::   :::||||||:::::::::
 456 ........LeuGluCysPheSerLysGlyTyrMetLeuSerGluGluPro 469

1486 ......CAGGCCCAGCAGCAGCAGCAGCAGCACCCGGTTTTACACCTTCA 1529
           :::|||:::   |||       :::   ||||||   |||   ||
 470 TyrIleHisAlaAsnTyrGlnProValGluIleProValSerHis...Gl 485

1530 GCCCCAGCAGATAATGCAGCTCCAGCAGCAGCAGCAGCAGCAGATCTCTC 1579
     ||||::::::          |||   :::::::::::::::::::::|||:
 485 nProGluSerLysAlaAlaLeuLeuLysLysLysAsnSerSerPheSerL 502

1580 AGCAACCTTACCCCCAGCAGCCGCCGCAT..................... 1608
     :::::   :::       :::      |||
 502 ysLysAspPheAlaProSerGluLysHisGluGlnAlaAspGluAspLeu 518

1608 ..................................................... 1608

 519 LeuSerGlnTyrGluAsnGlySerSerThrValValGluAlaLysThrSe 535

1609 .........CCATTTTCACAGCAGCAGCAGCAGCAA.............. 1635
             |||||:::::::::   :::   :::
 535 rGluAlaArgProPheAsnAspSerThrHisAlaGluProLeuAsnAspS 552

1636 ................CAGCAGCCACCACCATCGCCTCAGCAGCATCAG 1668
                     |||:::       |||      :::|||
 552 erThrHisIleSerLeuGlnGluAsnGlnSerSerValSerHisCys 568

1669 CTTTTTGGACATGATCCAGCAGTGGAG...ATTCCAGAAGAAGGCTTCTT 1715
     :::            |||   |||:::   |||   ||||||||||:::::
 569 Val............ProAspValSerThrIleThrGluGluGlyLeuPh 581

1716 ATTGGGATGTGTGTTTGCAATTGCCGATTATCCAGAGCAGATGTCTGATA 1765
     :           |||   :::       :::   ::::::   :::

 581 eSerGlnLysSerPheLeuValLeuGlyPheSerAsnGluAsnGlu.... 596

1766 AGCAACTGCTGGCCACCTGGAAAAGGATAATCCAGGCACATGGCGGCACT 1815
                :::::       :::|||||||:::   ::::::|||
 597 ..........SerAsnIleAlaAsnIleIleLysGluAsnAlaGlyLys 609

1816 GTTGACCCCACCTTCACGAGTCGATGCACGCAC.................. 1848
     :::       ::::::   ||||||   :::
 610 Ile...MetSerLeuLeuSerArgThrValAlaAspTyrAlaValValPr 625

1849 ....CTTCTCTGTGAGAGTCAAGTCAGCAGCGCGTATGCACAGGCAATAA 1894
         |||   ||||||   :::::::::   :::
 625 oLeuLeuGlyCysGluValGluAlaThrValGly................ 636

1895 GAGAAAGAAAGAGATGTGTTACTGCACACTGGTTAAACACAGTCTTAAAA 1944
                :::   ||||||       ||||||   |||   :::
 637 ..........GluValValThrAsnThrTrpLeuValThrCysIleAsp 649

1945 AAAAAAAAAATGGTA...CCGCCGCACCGAGCCCTTCACTTCCCAGTGGC 1991
     :::   :::       |||       :::   |||       ||||||
 650 TyrGlnThrLeuPheAspProLysSerAsnProLeuPheThrProValPr 666

1992 CTTCCCACCAGGAGGAAAGCCATGTTCACAGCATATTATTTCTGTGACTG 2041
                  |||     |||   ::::::   :::|||||   :::
 666 oValMetThrGlyMetThrProLeuGluAspCysValIleSerPheSerG 683

2042 GATTTGTTGATAGTGACAGAGATGACCTAAAATTAATGGCTTATTTGGCA 2091
         :::   :::::::::::::::|||   ::::::|||   |||
 683 lnCysAlaGlyAlaGluLysGluSerLeuThrPheLeuAlaAsnLeuLeu 699

2092 GGTGCCAAATATACGGGTTATCTATGCCGC..................... 2121
     |||||||:::          |||:::   |||
 700 GlyAlaSerValGlnGluTyrPheValArgLysSerAsnAlaLysLysGl 716

2122 .......AGCAACACAGTCCTCATCTGTAAAGAACCAACTGGTTTAAAGT 2164
            :::::::|||   ||||||   ||||||       |||   ||||
 716 yMetPheAlaSerThrHisLeuIleLeuLysGluArgGlyGlySerLysT 733

2165 ATGAAAAAGCCAAAGAGTGGAGGATACCCTGTGTCAACGCCCAGTGGCTT 2214
     |||||   ||||||:::|||::::::|||:::|||:::      ||||||
 733 yrGluAlaAlaLysLysTrpAsnLeuProAlaValThrIleAlaTrpLeu 749

2215 ..................................GGCGACATTCTTCT 2228
                                        :::   :::|||::
 750 LeuGluThrAlaArgThrGlyLysArgAlaAspGluSerHisPheLeuIl 766

2229 GGGAAAC..................................... 2235
     :   |||
 766 eGluAsnSerThrLysGluGluArgSerLeuGluThrGluIleThrAsnG 783

2235 ..................................................... 2235

 783 lyIleAsnLeuAsnSerAspThrAlaGluHisProGlyThrArgLeuGln 799

2236 .....................................................TT 2237

 800 ThrHisArgLysThrValValThrProLeuAspMetAsnArgPheGlnSe 816

2238 TGAGGCACTGAGGCAGATT......CAGTATAGTCGCTACACGGCATTCA 2281
     :::|||:::|||   :::      |||:::::|||   :::|||   |
 816 rLysAlaPheArgAlaValValSerGlnHisAlaArgGlnValAlaAlaS 833

2282 GT..................................................... 2283
     ||
 833 erProAlaValGlyGlnProLeuGlnLysGluProSerLeuHisLeuAsp 849

2283 ..................................................... 2283

 850 ThrProSerLysPheLeuSerLysAspLysLeuPheLysProSerPheAs 866

2284 .....................CTGCAGGATCCATTTGCCCCTACCCAGC 2311
                          |||:::   |||      |||:::|||:
 866 pValLysAspAlaLeuAlaAlaLeuGluThrProGlyArgProSerGlnG 883
```

```
 2312 AT..............................TTAGTTTTAAATCTT 2328
      ::                             :::|||  ||||||
  883 lnLysArgLysProSerThrProLeuSerGluValIleValLysAsnLeu 899

 2329 TTAGATGCTTGGAGAGTTCCCTTAAAAGTGTCTGCAGAGTTGTTGATGAG 2378
      :::::  |||     |||::::::     :::::
  900 ............GlnLeuAlaLeuAlaAsnSerSerArgAsnAlaValAl 912

 2379 TATAAGACTACCTCCCAAACTGAAACAGAATGAAGTAGCTAATGTCCAGC 2428
      ::::       |||:::||||||:::
  912 aLeuSerAlaSerProGlnLeuLysGlu..................... 921

 2429 CTTCTTCCCAAAGAGCCAGAATTGAAGACGTACCACCTCCCACTAAAAAG 2478
           :::::::::  |||::::::|||  |||     |||
  922 ........AlaGlnSerGluLysGluGluAlaProLysProLeuHisLys 935

 2479 CTAACTCCAGAATTGACCCCTTTTGTGCTTTTCACTGGATTCGAGCCTGT 2528
      ::::::     ::::::
  936 ValValValCysValSer.............................. 941

 2529 CCAGCTTCAACAGTATATTAAGAAGCTCTACATTCTTGGTGGAGAGGTTG 2578
           ::::::     ||||||         |||:::|||  :::|
  942 .......LysLysLeuSerLysLysGlnSerGluLeuAsnGlyIleAlaA 956

 2579 CGGAGTCTGCA.....................CAGAAGTGCACACAC 2604
      ||:::  :::                      :::     ||||||
  956 laSerLeuGlyAlaAspTyrArgTrpSerPheAspGluThrValThrHis 972

 2605 CTCATT............GCCAGCAAAGTGACTCGCACCGTGAAGTTCCT 2642
      :::|||             :::  :::::::|||     |||  ::
  973 PheIleTyrGlnGlyArgProAsnAspThrAsnArgGluTyrLysSerVa 989

 2643 GACGGCGATTTCTGTCGTGAAGCACATAGTCACGCCAGAGTGGCTGGAAG 2692
      :        :::|||     ||||||||||:::  :::||||||  :
  989 lLysGluArgGlyVal......HisIleValSerGluHisTrpLeuLeuA 1004

 2693 AATGCTTCAGGTGTCAGAAGTTCATTGATGAGCAGAACTAC 2733
      ::|||  :::     |||  :::  |||:::  |||
 1004 spCysAlaGlnGluCysLysHisLeuProGluSerLeuTyr 1017
```

seq_name: sp_fungi:Q9P5J5

seq_documentation_block:

```
ID   Q9P5J5      PRELIMINARY;      PRT;   838 AA.
AC   Q9P5J5;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CONSERVED HYPOTHETICAL PROTEIN.
GN   B23L21.340.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5141;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA   Nyakatura G., Mewes H.W., Mannhaupt G.;
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   German Neurospora genome project;
RL   Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL356172; CAB91702.2; -.
KW   Hypothetical protein.
SQ   SEQUENCE   838 AA;  93662 MW;  F75581407D6823D7 CRC64;
```

alignment_scores:

```
            Quality:  229.00            Length:    725
              Ratio:  0.727               Gaps:     21
Percent Similarity:  43.448   Percent Identity: 18.207
```

alignment_block:

US-09-664-641-10 x Q9P5J5 ..

Align seg 1/1 to: Q9P5J5 from: 1 to: 838

```
 1156 GTCCCACCCGTCCCAGGTAACATTTTGCCCCCTGAGGTCCGGGGTAATTT 1205
      :::|||  |||  :::|||  :::        :::      |||
   78 IleProValValAsnSerAsnTrpIleLysIleThrLeuAlaArgAsnLy 94

 1206 AATGGCTGCTGGACAAAACCTCCAAAGTTCTGAAAGATCAGAAATGATAG 1255
       :::|||                                  ::::::
   94 sValAla.....................................GlnValA 99

 1256 CTACCTGGAGTCCAGCTGTACGGACACTGAGGAAT......ATTACTAAT 1299
         :::||||||      |||  :::  :::      |||  :::
   99 rgProTyrSerProAspProArgMetIlePheSerAsnValIleLeuThr 115

 1300 AATGCTGACATTCAGCAGATGAACCGGCCATCAAATGTAGCACATATCTT 1349
       |||||||||  :::  ::::::  :::  ::::::      ::
  116 CysAlaAspIleProGluSerAspLysGluThrIleIleGlyAlaThrMe 132

 1350 ACAGACTCTTTCAGCACCTACGAAAAATTTAGAACAGCAGGTGAATCACA 1399
      :  :::      :::|||:::|||  ::::|||::::::|||
  132 tAlaLeuGlyGlyMetGluSerLysAspLeuThrArgGlnThrThrHisI 149

 1400 GCCAGCAGGGACATACAAATGCCAATGCAGTGCTGTTTAGCCAAGTGAAA 1449
  149 le..................................................... 149

 1450 GTGACTCCAGAGACACACATGCTACAGCAGCAGCAGCAGGCCCAGCAGCA 1499
           :::   ::::::::::::::::::::|||:::
  150 ............CysAlaLeuSerMetAspHisGluLysCysGlnGluAl 162

 1500 GCAGCAGCAGCACCCGGTTTTACACCTTCAGCCCCAGCAGATAATGCAGC 1549
       |||:::::::::|||         :::         :::   ::::::
  162 aGlnLysLysAsnProLysCysLysIleValLeuProHisTrpPheAspA 179

 1550 TCCAGCAGCAGCAGCAGCAGCAGATCTCTCAGCAACCTTACCCCCAGCAG 1599
            :::      ::::::|||||||:::   ||||||
  179 spCysPheArgLeuGlyArgArgIleSerGluAlaProTyrMetLeu... 194

 1600 CCGCCGCATCCATTTTCACAGCAGCAGCAGCAGCAACAGCAGCCACCACC 1649
                                                     ||
  195 ..................................................Pr 195

 1650 ATCGCCTCAGCAGCATCAGCTTTTTGGACATGATCCAGCAGTGGAGATTC 1699
      |:::|||:::      :::      |||  |||      :::  ||||
  195 oAsnProGluIleLeuArgThr...GlyProAspGluAspIleThrIleP 211

 1700 CAGAAGAAGGCTTCTTATTGGGATGTGTGTTTGCAATT......GCGGAT 1743
      ||  :::      :::  |||:::::::  :::|||      |||
  211 roProSerGluAlaValGluGlyAlaThrSerValIleProSerAlaAla 227

 1744 TATCCAGAGCAGATGTCTGATAAGCAACTGCTGGCCACCTGGAAAAGGAT 1793
       |||  :::      :::|||  ::::::  :::  |||:::::
  228 MetProGlyArgGluArgGluLysIleValValPheAlaGlnLysLysVa 244

 1794 AATC..................................................... 1797
      ::::
  244 lMetIleSerGlnAspLeuProIleAsnAlaArgLeuArgAsnIleLeuA 261

 1798 ..............CAGGCACATGGCGGCACTGTTGACCCCACCTTCACG 1833
                   |||:::  |||  :::|||:::         :::
  261 snAsnLeuIleGluGlnSerAspGlyGluValValAsnAsp......Val 275

 1834 AGTCGATGCACGCACCTTCTCTGTGAGAGTCAAGTCAGCAGCGCGTATGC 1883
      :::  |||          :::|||:::  :::  ::::::  |||
  276 AspAlaCysAspMetTyrValCysGlnTyrArgAspGlyAspGluTyrIl 292

 1884 ACAGGCAATAAGAGAAAGAAAGAGATGTGTTACTGCACACTGGTTAAACA 1933
       :::|||  ::::::  |||        :::      ||||||
  292 eArgAlaAlaGlnGlnGlyLysAspValGlyAsnLeuAlaTrpLeuTyrT 309
```

```
1934 CAGTCTTAAAAAAAAAAAAAATGGTA...CCGCCGCACCGAGCCCTTCAC 1980
     ::::::      ::::::   :::   |||   :::|||   ||||||
 309 yrLeuIleValHisAsnGluTrpThrArgProThrArgArgLeuLeuHis 325

1981 TTCCCAGTGGCCTTCCCACCAGGAGGAAAGCCATGTTCACAGCATATT.. 2028
     :::|||:::      |||      |||   |||            :::
 326 TyrProIle......ProArgAspGlyIleProGlyPheThrGlyMetLy 340

2029 .ATTTCTGTGACTGGATTTGTTGATAGTGACAGAGATGACCTAAAATTAA 2077
      |||:::::::::::::::      :::   |||      |||:::   :
 340 sIleThrLeuSerAsnTyrGlyGlyAspAlaArgValTyrLeuGluAsnL 357

2078 TGGCTTATTTGGCAGGTGCCAAATATACGGGTTATCTATGCCGCAGCAAC 2127
     ::        |||||||||   ||||||      :::      :::|||
 357 euIleThrAlaAlaGlyAlaThrTyrThrLysThrMetLysAlaAspAsn 373

2128 ACAGTCCTCATCTGTAAAGAACCAACTGGTTTAAAGTATGAAAAAGCCAA 2177
     |||   ||||||      :::   :::|||   |||   |||   |||||
 374 ThrHisLeuIleThrAlaArgMetAsnGlyGluLysCysGluAlaAlaLy 390

2178 ACAGTGGAGGATACCCTGTGTCAACGCCCAGTGG..............  2211
     |||||||:::|||      :::|||      ||
 390 sGluTrpAsnIleGluIleIleAsnHisLeuTrpIleGluAspSerTyrA 407

2211 ..................................................  2211

 407 laLysCysGluValGlnThrPheSerAsnProArgTyrGlnHisPhePro 423

2212 ............CTTGGCGACATTCTTCTGGGAAACTTT.........GA 2240
                 ||||||:::|||:::   ::::::|||         ::
 424 ProArgThrAsnLeuGlyGluIleIleGlySerThrPhePheAspGluGl 440

2241 GGCACTGAGGCAGATTCAGTAT.......................... 2262
     :   ||||||::::::   :::
 440 nArgLeuArgGluValTyrPheProGlyGlyGluGluValLeuSerThrA 457

2263 .....................................AGTCGCTACACG 2274
                                         ::::::
 457 laAlaLysLysArgArgLysAlaAsnAspAlaAlaAsnLysAsnValTyr 473

2275 GCATTCAGTCTGCAGGATCCATTTGCCCCTACCCAGCATTTAGTTTTAAA 2324
     |||   :::|||   :::      |||   |||   :::
 474 AlaSerAsnLeuAlaGluThrAspAlaGlyThrAlaAsnLysSerSerPr 490

2325 TCTTTTAGATGCTTGGAGAGTTCCCTTAAAAGTGTCTGCAGAGTTGTTGA 2374
      |||            :::            :::|||:::      :
 490 oLeuGlyGlyLysThrAlaIleLysProProArgGlyAlaGlnAsnLysG 507

2375 TGAGTATAAGACTACCTCCCAAACTG.................AAACAG 2406
     :::::|||:::   |||   |||                    |||:::
 507 lnAsnIleGlnThrProAlaLysThrThrArgThrThrArgGlyLysGlu 523

2407 AATGAAGTAGCTAATGTCCAGCCTTCTTCCCAAAGA.............. 2442
     |||::::::   :::|||:::   |||:::::|||
 524 AsnAspThrProSerValMetSerSerGlySerArgSerAlaLysAspLy 540

2443 .......GCCAGAATTGAAGACGTACCACCT.................C 2467
             ::::::::::::::::::   |||
 540 sAlaLeuSerLysLeuHisGlnLeuAlaProAspIleAlaLeuTyrGluL 557

2468 CCACTAAAAAGCTAACTCCAGAATTGACCCCTTTTGTGCTTTTCACTGGA 2517
       ||||||   |||   :::      |||:::            :::
 557 ysGluLysLysArgThrAlaLysAspGlyProTrpGlyGlyLysArgAla 573

2518 TTCGAGCCTGTCCAGGTTCAACAGTATATTAAGAAGCTCTACATTCTT.. 2565
        :::   ::::::   ::::::      |||:::
 574 AlaAspGlnIleAspArgAspArgAlaAlaLysGluSerSerProAlaGl 590

2566 ....GGTGGAGAGGTTGCGGAGTCTGCACAGAAG.............. 2595
       |||||||||      |||:::   ||||||
 590 uAspGlyGlyGluGluAspGluThrGluGlnLysArgProAlaLysLysA 607

2595 ..................................................  2595

 607 laArgValSerSerSerProGluMetArgIleCysLeuThrGlyTyrLys 623

2595 ..................................................  2595

 624 ArgTrpValAsnAspLysMetArgGluGluAlaAspArgIleValGlnAs 640

2596 ..........TGCACACACCTCATTGCCAGCAAAGTGACTCGCACCGTGA 2635
                |||   :::|||   |||   :::::::::|||||||:::|
 640 pAsnValProCysAspTyrLeuAlaAlaProArgMetValArgThrMetL 657

2636 AGTTCCTGACGGCGATTTCTGTCGTGAAGCACATAGTGACGCCAGAGTGG 2685
     ||||||||   :::::::::            |||::::::   ::::::
 657 ysPheLeuArgCysLeuAlaArgGlyProAspIleIleSerSerAspTyr 673

2686 CTGGAAGAATGCTTCAGGTGTCAGAAGTTCATTGATGAGCAGAACTACAT 2735
     :::      |||   ::::::   |||   :::      :::::::|||::
 674 ValThrAlaCysValGluAlaGlyLysValLeuProProLysAspTyrLe 690

2736 TCTCCGAGATGCTGAGGCAGAAGTACTTTTCTCTTTCAGCTTGGAAGAAT 2785
     :|||   |||   |||:::|||      |||:::   :::|||:::   :
 690 uLeuValAspLysGluSerGluAspArgPheGlyValThrLeuGlnThrA 707

2786 CCTTAAAACGGGCACACGTTTCTCCACTCTTTAAGGCAAAATATTTTTAC 2835
     ::::::::||||||:::::::::            :::::::::   ::::::
 707 laIleSerArgAlaArgAlaAsn.........ArgGlyArgLeuLeuTrp 720

2836 ATCACACCTGGAATCTGCCCA..............AGTCTTTCCACTAT 2870
       ::::||   :::|||                  ::::::::::::
 721 GlyValProValPheCysThrGluGluIleLysAsnGlyValGlnSerTy 737

2871 GAAGGCAATCGTAGAGTGTGCAGGA 2895
     ::::::|||:::|||:::   |||
 737 rGlnThrIleAlaGluAlaAsnGly 745
```

fragments

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 4, 2002, 11:09:22 ; Search time 6332.45 Seconds
(without alignments)
11830.650 Million cell updates/sec

Title: US-09-664-641-10
Perfect score: 3580
Sequence: 1 catagacagtcgatacgcct.....gaacttagaaaaaaaaaaaa 3580

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2664	74.4	2732	9	HSU80735	U80735 Homo sapien
2	1495	41.8	1518	9	BC008328	BC008328 Homo sapi
3	1483	41.4	3882	10	AF104261	AF104261 Mus muscu
4	874	24.4	3771	5	AF172855	AF172855 Xenopus l
5	636.6	17.8	149592	2	AC073077	AC073077 Homo sapi
c 6	636.6	17.8	159041	9	AC093726	AC093726 Homo sapi
c 7	479.2	13.4	149592	2	AC073077	AC073077 Homo sapi
c 8	361.4	10.1	438	11	G30194	G30194 human STS S
c 9	353.4	9.9	120887	2	AC094546	AC094546 Rattus no
10	348.6	9.7	178419	2	AC074040	AC074040 Mus muscu
11	334.4	9.3	177205	2	AC079217	AC079217 Mus muscu
c 12	308.2	8.6	633	9	HSU23862	U23862 Human clone
c 13	263.4	7.4	687	11	G15998	G15998 human STS C
14	204.6	5.7	690	6	AX186240	AX186240 Sequence
15	186.8	5.2	504	6	AX187445	AX187445 Sequence
16	186.8	5.2	579	6	AX184577	AX184577 Sequence
c 17	174.4	4.9	177205	2	AC079217	AC079217 Mus muscu
c 18	174.4	4.9	178419	2	AC074040	AC074040 Mus muscu
c 19	142.2	4.0	2122	9	AK056609	AK056609 Homo sapi
20	136	3.8	267	11	G24669	G24669 human STS W
21	136	3.8	462	11	G60373	G60373 SHGC-131148
c 22	135	3.8	209	9	U92824	U92824 Homo sapien
23	130.6	3.6	123554	9	AB023049	AB023049 Homo sapi
c 24	130.6	3.6	151228	9	AL662800	AL662800 Human DNA
25	130.6	3.6	200000	9	AP000512	AP000512 Homo sapi
26	130.6	3.6	221059	2	AL662797	AL662797 Homo sapi
27	123.2	3.4	444	11	G33227	G33227 human STS S
28	110.6	3.1	397	6	BD010479	BD010479 Diagnosti
29	105.6	2.9	5120	6	AX067148	AX067148 Sequence
30	105.6	2.9	5120	9	AF152102	AF152102 Homo sapi
31	105.2	2.9	1159	9	AF126748	AF126748 Homo sapi
c 32	105.2	2.9	1442	9	AF087653	AF087653 Homo sapi
c 33	105.2	2.9	1472	9	AF126749	AF126749 Homo sapi
34	104.8	2.9	720	9	AB007820	AB007820 Homo sapi
35	104.6	2.9	107637	2	AC094679	AC094679 Rattus no
36	104.6	2.9	177298	2	AC068751	AC068751 Homo sapi
37	103	2.9	234	6	AR002264	AR002264 Sequence
38	103	2.9	234	6	AR053098	AR053098 Sequence
39	100.6	2.8	119526	2	AC007468	AC007468 Drosophil
c 40	100.6	2.8	138685	2	AC017582	AC017582 Drosophil
41	100.6	2.8	166244	3	AC093502	AC093502 Drosophil
42	100.6	2.8	173851	2	AC096966	AC096966 Rattus no
c 43	100.6	2.8	185400	3	AC104703	AC104703 Drosophil
c 44	100.6	2.8	341319	3	AE003536	AE003536 Drosophil
45	100.2	2.8	477	6	AR165922	AR165922 Sequence

ALIGNMENTS

RESULT 1
HSU80735
LOCUS HSU80735 2732 bp mRNA linear PRI 18-DEC-1997
DEFINITION Homo sapiens CAGF28 mRNA, partial cds.
ACCESSION U80735
VERSION U80735.1 GI:2565045
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2732)
AUTHORS Margolis,R.L., Abraham,M.R., Gatchell,S.B., Li,S.H., Kidwai,A.S., Breschel,T.S., Stine,O.C., Callahan,C., McInnis,M.G. and Ross,C.A.
TITLE cDNAs with long CAG trinucleotide repeats from human brain
JOURNAL Hum. Genet. 100 (1), 114-122 (1997)
MEDLINE 97369492
PUBMED 9225980
REFERENCE 2 (bases 1 to 2732)
AUTHORS Margolis,R.L., Abraham,M.R., Gatchell,S.B., Li,S.H., Kidwai,A.S., Breschel,T.S., Stine,O.C., Callahan,C., McInnis,M.G. and Ross,C.A.
TITLE Direct Submission
JOURNAL Submitted (02-DEC-1996) Psychiatry, Johns Hopkins Univ. Sch. of

```
            Med., 600 N. Wolfe Street, Meyer 2-181, Baltimore, MD 21205, USA
COMMENT     Contains a polymorphic region having an allele frequency of 145:
            1.00 (allele bp). In 32 chromosomes tested, heterozygosity was 0%.
FEATURES             Location/Qualifiers
     source          1. .2732
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q; distal to WI-8540"
                     /tissue_type="brain"
                     /dev_stage="fetus"
                     /note="single strand sequence"
     CDS             <1. .2236
                     /note="contains glutamine repeats"
                     /codon_start=2
                     /product="CAGF28"
                     /protein_id="AAB91434.1"
                     /db_xref="GI:2565046"
                     /translation="ENEEQDSQNEGSTDEKSSPASSQEGSPSGDQQFSPKSNTEKSKG
                     ELMFDDSSDSSPEKQERNLNWTPAEVPQLAAAKRRLPQGKEPGLINLCANVPPVPGNI
                     LPPEVRGNLMAAGQNLQSSERSEMIATWSPAVRTLRNITNNADIQQMNRPSNVAHILQ
                     TLSAPTKNLEQQVNHSQQGHTNANAVLFSQVKVTPETHMLQQQQQAQQQQQQHPVLHL
                     QPQQIMQLQQQQQQQISQQPYPQQPPHPFSQQQQQQQQPPPSPQQHQLFGHDPAVEIP
                     EEGFLLGCVFAIADYPEQMSDKQLLATWKRIIQAHGGTVDPPSRVDARTFSVRVKSAA
                     RIAQAIREKKRCVTAHWLNTVLKKKKMVPPHRALHFPVAFPPGGKPCSQHIISVTGFV
                     DSDRDDLKLMAYLAGAKYTGYLCRSNTVLICKEPTGLKYEKAKEWRIPCVNAQWLGDI
                     LLGNFEALRQIQYSRYTAFSLQDPFAPTQHLVLNLLDAWRVPLKVSAELLMSIRLPPK
                     LKQNEVANVQPSSQRARIEDVPPPTKKLTPELTPFVLFTGFEPVQVQQYIKKLYILGG
                     EVAESAQKCTHLIASKVTRTLKFLAAISVVKHIVTPEWLEECFRCQKFIDEQNYILRD
                     AEAEVLFSFSLEESLKRAHVSPLFKAKYFYITPGICPSLSTMKAIVECAGGKVLSKQP
                     SFRKLMEHKQNSSLSETILISCENDLHLCREYFARGIDVHNAEFVLTGVLTQTLDYES
                     YKFN"
     repeat_region   602. .616
                     /rpt_type=tandem
                     /rpt_unit=CAG
     repeat_region   620. .637
                     /rpt_type=tandem
                     /rpt_unit=CAG
     repeat_region   680. .700
                     /rpt_type=tandem
                     /rpt_unit=CAG
BASE COUNT      816 a    612 c    605 g    695 t      4 others
ORIGIN


  Query Match           74.4%;  Score 2664;  DB 9;  Length 2732;
  Best Local Similarity 99.4%;  Pred. No. 0;
  Matches 2692;  Conservative    0;  Mismatches   14;  Indels    2;  Gaps    2;

Qy    874 gaaaatgaggaacaagattctcagaatgagggtagtacagatgagaagtcaagccctgcc 933
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2 GAAAATGAGGAACAAGATTCTCAGAATGAGGGTAGTACAGATGAGAAGTCAAGCCCTGCC 61

Qy    934 agctctcaagaagggtctccttcaggtgaccagcagttttcacctaaatccaacactgaa 993
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     62 AGCTCTCAACAAGGGTCTCCTTCAGGTGACCAGCAGTTTTCACCTAAATCCAACACTGAA 121

Qy    994 aaatctaaaggggaattaatgtttgatgattcttcagattcatcaccggaaaaacaggag 1053
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122 AAATCTAAAGGGGAATTAATGTTTGATGATTCTTCAGATTCATCACCGGAAAAACAGGAG 181

Qy   1054 agaaatttaaactggaccccggccgaagtcccacagttagctgcagcaaaacgcaggctg 1113
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    182 AGAAATTTAAACTGGACCCCGGCCGAAGTCCCAAGTTAGCTGCAGCAAAACGCAGGCTG 241

Qy   1114 cctcagggaaaggagcctgggttgattaacttgtgtgccaatgtcccacccgtcccaggt 1173
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    242 CCTCAGGGAAAGGAGCCTGGGTTGATTAACTTGTGTGCCAATGTCCCACCCGTCCCAGGT 301

Qy   1174 aacattttgcccccctgaggtccggggtaatttaatggctgctggacaaaacctccaaagt 1233
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    302 AACATTTTGCCCCCTGAGGTCCGGGGTAATTTAATGGCTGCTGGACAAAACCTCCAAAGT 361

Qy   1234 tctgaaagatcagaaatgatagctacctggagtccagctgtacggacactgaggaatatt 1293
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    362 TCTGAAAGATCAGAAATGATAGCTACCTGGAGTCCAGCTGTACGGACACTGAGGAATATT 421

Qy   1294 actaataatgctgacattcagcagatgaaccggccatcaaatgtagcacatatcttacag 1353
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    422 ACTAATAATGCTGACATTCAGCAGATGAACCGGCCATCAAATGTAGCACATATCTTACAG 481

Qy   1354 actctttcagcacctacgaaaaatttagaacagcaggtgaatcacagccagcagggacat 1413
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    482 ACTCTTTCAGCACCTACGAAAAATTTAGAACAGCAGGTGAATCACAGCCAGCAGGGACAT 541

Qy   1414 acaaatgccaatgcagtgctgtttagccaagtgaaagtgactccagagacacacatgcta 1473
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    542 ACAAATGCCAATGCAGTGCTGTTTAGCCAAGTGAAAGTGACTCCAGAGACACACATGCTA 601

Qy   1474 cagcagcagcagcaggcccagcagcagcagcagcagcacccggttttacaccttcagccc 1533
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    602 CAGCAGCAGCAGCAGGCCCAGCAGCAGCAGCAGCACCCGGTTTTACACCTTCAGCCC 661

Qy   1534 cagcagataatgcagctccagcagcagcagcagcagcagatctctcagcaaccttacccc 1593
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    662 CAGCAGATAATGCAGCTCCAGCAGCAGCAGCAGCAGCAGATCTCTCAGCAACCTTACCCC 721

Qy   1594 cagcagccgccgcatccatttttcacagcagcagcagcagcaacagcagccaccaccatcg 1653
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    722 CAGCAGCCGCCGCATCCATTTTCACAGCAGCAGCAGCAGCAACAGCAGCCACCACCATCG 781

Qy   1654 cctcagcagcatcagctttttggacatgatccagcagtggagattccagaagaaggctic 1713
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    782 CCTCAGCAGCATCAGCTTTTTGGACATGATCCAGCAGTGGAGATTCCAGAAGAAGGCTTC 841

Qy   1714 ttattgggatgtgtgtttgcaattgcggattatccagagcagatgtctgataagcaactg 1773
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    842 TTATTGGGATGTGTGTTTGCAATTGCGGATTATCCAGAGCAGATGTCTGATAAGCAACTG 901

Qy   1774 ctggccacctggaaaaggataatccaggcacatggcggcactgttgaccccacccttcacg 1833
          |||||||||||||||||||||||||||||||||||||||| ||||||| ||||||||||||
Db    902 CTGGCCACCTGGAAAAGGATAATCCAGGCACATGGCGGNACTGTTGA-CCCACCTTCACG 960

Qy   1834 agtcgatgcacgcaccttctctgtgagagtcaagtcagcagcgcgta-tgcacaggcaat 1892
          |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db    961 AGTCGATGCACGCACCTTCTCTGTGAGAGTCAAGTCAGCAGCGCGTATTGCACAGGCAAT 1020

Qy   1893 aagagaaagaaagagatgtgttactgcacactggttaaacacagtcttaaaaaaaaaaaa 1952
          |||||||||||||||||||||||||||||||||||||||||||||||||||| || || ||
Db   1021 AAGAGAAAGAAAGAGATGTGTTACTGCACACTGGTTAAACACAGTCTTAAAGAAGAAGAA 1080

Qy   1953 aatggtaccgccgcaccgagcccttcacttcccagtggccttcccaccaggaggaaagcc 2012
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 AATGGTACCGCCGCACCGAGCCCTTCACTTCCCAGTGGCCTTCCCACCAGGAGGAAAGCC 1140

Qy   2013 atgttcacagcatattatttctgtgactggatttgttgatagtgacagagatgacctaaa 2072
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 ATGTTCACAGCATATTATTTCTGTGACTGGATTTGTTGATAGTGACAGAGATGACCTAAA 1200

Qy   2073 attaatggcttatttggcaggtgccaaatatacgggttatctatgccgcagcaacacagt 2132
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 ATTAATGGCTTATTTGGCAGGTGCCAAATATACGGGTTATCTATGCCGCAGCAACACAGT 1260

Qy   2133 cctcatctgtaaagaaccaactggtttaaagtatgaaaaagccaaagagtggaggatacc 2192
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 CCTCATCTGTAAAGAACCAACTGGTTTAAAGTATGAAAAAGCCAAAGAGTGGAGGATACC 1320

Qy   2193 ctgtgtcaacgcccagtggcttggcgacattcttctgggaaactttgaggcactgaggca 2252
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 CTGTGTCAACGCCCAGTGGCTTGGCGACATTCTTCTGGGAAACTTTGAGGCACTGAGGCA 1380

Qy   2253 gattcagtatagtcgctacacggcattcagtctgcaggatccatttgcccctacccagca 2312
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 GATTCAGTATAGTCGCTACACGGCATTCAGTCTGCAGGATCCATTTGCCCCTACCCAGCA 1440

Qy   2313 tttagttttaaatcttttagatgcttggagagttcccttaaaagtgtctgcagagttgtt 2372
```

```
Db   1441 TTTAGTTTTAAATCTTTTAGATGCTTGGAGAGTTCCCTTAAAAGTGTCTGCAGAGTTGTT 1500

Qy   2373 gatgagtataagactacctcccaaactgaaacagaatgaagtagctaatgtccagccttc 2432
Db   1501 GATGAGTATAAGACTACCTCCCAAACTGAAACAGAATGAAGTAGCTAATGTCCAGCCTTC 1560

Qy   2433 ttcccaaagaqccagaattgaagacgtaccacctcccactaaaaagctaactccagaatt 2492
Db   1561 TTCCCAAAGAGCCAGAATTGAAGACGTACCACCTCCCACTAAAAAGCTAACTCCAGAATT 1620

Qy   2493 gaccccttttgtgctttcactqgattcgagcctgtccaggttcaacagtatattaagaa 2552
Db   1621 GACCCCTTTTGTGCTTTTCACTGGATTCGAGCCTGTCCAGGTTCAACAGTATATTAAGAA 1680

Qy   2553 gctctacattcttggtggagaggttgcggagtctgcacagaagtgcacacacctcattgc 2612
Db   1681 GCTCTACATTCTTGGTGGAGAGGTTGCGGAGTCTGCACAGAAGTGCACACACCTCATTGC 1740

Qy   2613 cagcaaagtgactcgcaccgtgaagttcctgacggcgatttctgtcgtgaagcacatagt 2672
Db   1741 CAGCAAAGTGACTCGCACCCTGAAGTTCCTGGCGGCGATTTCTGTCGTGAAGCACATAGT 1800

Qy   2673 gacgccagagtqgctggaagaatgcttcaggtgtcagaagttcattgatgagcagaacta 2732
Db   1801 GACGCCAGAGTGGCTGGAAGAATGCTTCAGGTGTCAGAAGTTCATTGATGAGCAGAACTA 1860

Qy   2733 cattctccgagatgctgaggcagaagtacttttctctttcagcttggaagaatccttaaa 2792
Db   1861 CATTCTCCGAGATGCTGAGGCGGAAGTACTTTTCTCTTTCAGCTTGGAAGAATCCTTAAA 1920

Qy   2793 acgggcacacgtttctccactctttaaggcaaaatattttacatcacacctggaatctg 2852
Db   1921 ACGGGCACACGTTTCTCCACTCTTTAAGGCAAAATATTTTTACATCACACCTGGAATCTG 1980

Qy   2853 cccaagtctttccactatgaaggcaatcgtagagtgtgcaggaggaaaggtgttatccaa 2912
Db   1981 CCCAAGTCTTTCCACTATGAAGGCAATCGTAGAGTGTGCAGGAGGAAAGGTGTTATCCAA 2040

Qy   2913 gcagccatctttccggaagctcatggagcacaagcagaactcgagtttgtcggaaataat 2972
Db   2041 GCAGCCATCTTTCCGGAAGCTCATGGAGCACAAGCAGAACTCGAGTTTGTCGGAAATAAT 2100

Qy   2973 tttaatatcctgtgaaaatgaccttcatttatgccgagaatattttgccagaggcataga 3032
Db   2101 TTTAATATCCTGTGAAAATGACCTTCATTTATGCCGAGAATATTTTGCCAGAGGCATAGA 2160

Qy   3033 tgttcacaatgcagagttcgttctgactggagtgctcactcaaacgctggactatgaatc 3092
Db   2161 TGTTCACAATGCAGAGTTCGTTCTGACTGGAGTGCTCACTCAAACGCTGGACTATGAATC 2220

Qy   3093 atataagtttaactgatggcgtctaggctgccgtgcatgtcgactcctgcggtgcggggc 3152
Db   2221 ATATAAGTTTAACTGATGGCGTCTAGGCTGCCGTGCATGTCGACTCCTGCGGTGCGGGGC 2280

Qy   3153 tggctgtctggctggcgaggagctgctgcgcttccttcacatgctcttgttttccagctg 3212
Db   2281 TGGCTGTCTGGCTGGCGAGGAGCTGCTGCGCTTCCTTCACGTGCTCTTGTTTTCCAGCTG 2340

Qy   3213 ctttcctggggatcagactgtgaagcaggaagacagatataataaatatactgcatctt 3272
Db   2341 CTTTCCTGGGGGATCAGACTGTGAAGCAGGAAGACAGATATAATAAATATACTGCATCTT 2400

Qy   3273 tttaagatgtgcaattttattctgaggaaacataaattatgttttgtattatatgacttt 3332
Db   2401 TTTAAGATGTGCAATTTTATTCTGAGGAAACATAAATTATGTTTTGTATTATATGACTTT 2460

Qy   3333 aagagcccacattaggttttatgattcattlgccaggtttttaaatgttttcacaaaact 3392
Db   2461 AAGAGCCCACATTAGGTTTTATGATTCATTTGCCAGGTTTTTAAATGTTTTCACAAAACT 2520

Qy   3393 gttacgggacttcaactagaaataaaatggtgtaaataaagaccttgctatctctaaatt 3452
Db   2521 GTTACGGGACTTCAACTAGAAATAAAATGGTGTAAATAAAGACCTTGCTATCTCTAAATT 2580

Qy   3453 atggatgttaaagatttgaaatgttttgtactttgattatttttatttcttatactctgt 3512
Db   2581 ATGGATGTTAAAGATTTGAAATGTTTTGTACTTTGATTATTTTTATTTCTAAAACTCTGT 2640

Qy   3513 tttcttttatattgatatcttgcccacattttaaataaatgtacttttgaactlagaaaa 3572
Db   2641 TTTCTTTTANATTGANATCTTGCCCACATTTTAAATAAATGTACTTTTGNACTTAAAAAA 2700

Qy   3573 aaaaaaaa 3580
Db   2701 AAAAAAAA 2708
```

```
RESULT   2
BC008328
LOCUS       BC008328               1518 bp    mRNA    linear   PRI 12-JUL-2001
DEFINITION  Homo sapiens, clone IMAGE:3503689, mRNA, partial cds.
ACCESSION   BC008328
VERSION     BC008328.1  GI:14249897
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1518)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAY-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
            Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy
            George Yang, Scott Zuyderduyn, Marco Marra.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 23 Row: c Column: 4.
FEATURES             Location/Qualifiers
     source          1. .1518
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3503689"
                     /tissue_type="Brain, neuroblastoma"
                     /clone_lib="NIH_MGC_19"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     CDS             <1. .1177
                     /codon_start=2
                     /product="Unknown (protein for IMAGE:3503689)"
                     /protein_id="AAH08328.1"
                     /db_xref="GI:14249898"
                     /translation="TVLKKKKMVPPHRALHFPVAFPPGGKPCSQHIISVTGFVDSDRD
                     DLKLMAYLAGAKYTGYLCRSNTVLICKEPTGLKYEKAKEWRIPCVNAQWLGDILLGNF
                     EALRQIQYSRYTAFSLQDPFAPTQHLVLNLLDAWRVPLKVSAELLMSIRLPPKLKQNE
                     VANVQPSSKRARIEDVPPPTKKLTPELTPFVLFTGFEPVQVQQYIKKLYILGGEVAES
                     AQKCTHLIASKVTRTVKFLTAISVVKHIVTPEWLEECFRCQKFIDEQNYILRDAEAEV
```

```
                    LFSFSLEESLKRAHVSPLFKAKYFYITPGICPSLSTMKAIVECAGGKVLSKQPSFRKL
                    MEHKQNSSLSEILISCENDLHLCREYFARGIDVHNAEFVLTGVLTQTLDYESYKFN"
BASE COUNT      439 a    322 c    339 g    418 t
ORIGIN
```

Query Match 41.8%; Score 1495; DB 9; Length 1518;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1498; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy   1932 cacagtcttaaaaaaaaaaaaaatggtaccgccgcaccgagcccttcacttcccagtggc 1991
          ||||||||||| || || ||||||||||||||||||||||||||||||||||||||||||
Db      1 CACAGTCTTAAAGAAGAAGAAAATGGTACCGCCGCACCGAGCCCTTCACTTCCCAGTGGC 60

Qy   1992 cttcccaccaggaggaaagccatgttcacagcatattatttctgtgactggatttgttga 2051
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CTTCCCACCAGGAGGAAAGCCATGTTCACAGCATATTATTTCTGTGACTGGATTTGTTGA 120

Qy   2052 tagtgacagagatgacctaaaattaatggcttatttggcaggtgccaaatatacgggtta 2111
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TAGTGACAGAGATGACCTAAAATTAATGGCTTATTTGGCAGGTGCCAAATATACGGGTTA 180

Qy   2112 tctatgccgcagcaacacagtcctcatctgtaaagaaccaactggtttaaagtatgaaaa 2171
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TCTATGCCGCAGCAACACAGTCCTCATCTGTAAAGAACCAACTGGTTTAAAGTATGAAAA 240

Qy   2172 agccaaagagtggaggataccctgtgtcaacgcccagtggcttggcgacattcttctggg 2231
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AGCCAAAGAGTGGAGGATACCCTGTGTCAACGCCCAGTGGCTTGGCGACATTCTTCTGGG 300

Qy   2232 aaactttgaggcactgaggcagattcagtatagtcgctacacggcattcagtctgcagga 2291
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AAACTTTGAGGCACTGAGGCAGATTCAGTATAGTCGCTACACGGCATTCAGTCTGCAGGA 360

Qy   2292 tccatttgcccctacccagcatttagttttaaatctttagatgcttggagagttccctt 2351
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TCCATTTGCCCCTACCCAGCATTTAGTTTTAAATCTTTAGATGCTTGGAGAGTTCCCTT 420

Qy   2352 aaaagtgtctgcagagttgttgatgagtataagactacctcccaaactgaaacagaatga 2411
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AAAAGTGTCTGCAGAGTTGTTGATGAGTATAAGACTACCTCCCAAACTGAAACAGAATGA 480

Qy   2412 agtagctaatgtccagccttcttcccaaagagccagaattgaagacgtaccacctcccac 2471
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db    481 AGTAGCTAATGTCCAGCCTTCTTCCAAAAGAGCCAGAATTGAAGACGTACCACCTCCCAC 540

Qy   2472 taaaaagctaactccagaattgacccctttgtgctttcactggattcgagcctgtcca 2531
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TAAAAAGCTAACTCCAGAATTGACCCCTTTTGTGCTTTCACTGGATTCGAGCCTGTCCA 600

Qy   2532 ggttcaacagtatattaagaagctctacattcttggtggagaggttgcggagtctgcaca 2591
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GGTTCAACAGTATATTAAGAAGCTCTACATTCTTGGTGGAGAGGTTGCGGAGTCTGCACA 660

Qy   2592 gaagtgcacacacctcattgccagcaaagtgactcgcaccgtgaagttcctgacggcgat 2651
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GAAGTGCACACACCTCATTGCCAGCAAAGTGACTCGCACCGTGAAGTTCCTGACGGCGAT 720

Qy   2652 ttctgtcgtgaagcacatagtgacgccagagtggctggaagaatgcttcaggtgtcagaa 2711
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 TTCTGTCGTGAAGCACATAGTGACGCCAGAGTGGCTGGAAGAATGCTTCAGGTGTCAGAA 780

Qy   2712 gttcattgatgagcagaactacattctccgagatgctgaggcagaagtacttttctcttt 2771
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GTTCATTGATGAGCAGAACTACATTCTCCGAGATGCTGAGGCAGAAGTACTTTTCTCTTT 840

Qy   2772 cagcttggaagaatccttaaaacgggcacacgtttctccactctttaaggcaaaatattt 2831
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 CAGCTTGGAAGAATCCTTAAAACGGGCACACGTTTCTCCACTCTTTAAGGCAAAATATTT 900

Qy   2832 ttacatcacacctggaatctgcccaagtctttccactatgaaggcaatcgtagagtgtgc 2891
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 TTACATCACACCTGGAATCTGCCCAAGTCTTTCCACTATGAAGGCAATCGTAGAGTGTGC 960

Qy   2892 aggaggaaaggtgttatccaagcagccatctttccggaagctcatggagcacaagcagaa 2951
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 AGGAGGAAAGGTGTTATCCAAGCAGCCATCTTTCCGGAAGCTCATGGAGCACAAGCAGAA 1020

Qy   2952 ctcgagtttgtcggaaataatttttaatatcctgtgaaaatgaccttcatttatgccgaga 3011
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 CTCGAGTTTGTCGGAAATAATTTTAATATCCTGTGAAAATGACCTTCATTTATGCCGAGA 1080

Qy   3012 atattttgccagaggcatagatgttcacaatgcagagttcgttctgactggagtgctcac 3071
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 ATATTTTGCCAGAGGCATAGATGTTCACAATGCAGAGTTCGTTCTGACTGGAGTGCTCAC 1140

Qy   3072 tcaaacgctggactatgaatcatataagtttaactgatggcgtctaggctgccgtgcatg 3131
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 TCAAACGCTGGACTATGAATCATATAAGTTTAACTGATGGCGTCTAGGCTGCCGTGCATG 1200

Qy   3132 tcgactcctgcggtgcggggctggctgtctggctggcgaggagctgctgcgcttccttca 3191
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 TCGACTCCTGCGGTGCGGGGCTGGCTGTCTGGCTGGCGAGGAGCTGCTGCGCTTCCTTCA 1260

Qy   3192 catgctcttgttttccagctgctttcctgggggatcagactgtgaagcaggaagacagat 3251
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 CATGCTCTTGTTTTCCAGCTGCTTTCCTGGGGGATCAGACTGTGAAGCAGGAAGACAGAT 1320

Qy   3252 ataataaatatactgcatctttttaagatgtgcaattttattctgaggaaacataaatta 3311
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 ATAATAAATATACTGCATCTTTTTAAGATGTGCAATTTTATTCTGAGGAAACATAAATTA 1380

Qy   3312 tgttttgtattatatgactttaagagcccacattaggttttatgattcatttgccaggtt 3371
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 TGTTTTGTATTATATGACTTTAAGAGCCCACATTAGGTTTTATGATTCATTTGCCAGGTT 1440

Qy   3372 tttaaatgttttcacaaaactgttacgggacttcaactagaaataaaatggtgtaaataa 3431
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 TTTAAATGTTTTCACAAAACTGTTACGGGACTTCAACTAGAAATAAAATGGTGTAAATAA 1500

Qy   3432 aga 3434
          ||| |
Db    1501 AAA 1503
```

```
RESULT   3
AF104261
LOCUS       AF104261               3882 bp    mRNA    linear   ROD 14-JUL-2000
DEFINITION  Mus musculus Pax transcription activation domain interacting
            protein PTIP mRNA, complete cds.
ACCESSION   AF104261
VERSION     AF104261.1  GI:4336733
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 3882)
  AUTHORS   Lechner,M.S., Levitan,I. and Dressler,G.R.
  TITLE     PTIP, a novel BRCT domain-containing protein interacts with Pax2
            and is associated with active chromatin
  JOURNAL   Nucleic Acids Res. 28 (14), 2741-2751 (2000)
  MEDLINE   20368635
REFERENCE   2  (bases 1 to 3882)
  AUTHORS   Lechner,M.S. and Dressler,G.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-1998) Pathology, Univ. of Michigan, 1150 W.
            medical Center Dr., Ann Arbor, MI 48109, USA
FEATURES             Location/Qualifiers
     source          1. .3882
                     /organism="Mus musculus"
                     /strain="FVB"
                     /db_xref="taxon:10090"
     CDS             272. .3442
                     /note="contains BRCT domains"
                     /codon_start=1
```

/product="Pax transcription activation domain interacting protein PTIP"
/protein_id="AAD17923.1"
/db_xref="GI:4336734"
/translation="MSEPAPEVPEELFREVKYYAVGDIDPQVIQLLKAGKAKEVSYNA
LASHIISEDGDNPEVGEAREVFDLPVVKPSWVTLSVQCGALLPVNGFSPESCQIFFGL
TACLSQVSSEDRSALWALVTFHGGSCQLNLNKKCTHLIVPEPKGEKYERAVKRTSIKI
VTPDWVLDCVSEKRRKDEAFYHPRLIIYEEEEEEEEEGDNEEQDSQNEGSTEKSSVAS
SAVASPAEQPCSPKPRAEVSKGELMFDDSSDSSPEKQERSLNWAPAEAPPLNTAQRRL
PQGKGPGLINLCANVPPVPGDILPPDMRGNLMAPGQNLQNSERSEILGTWSPAVRTLR
NITNNADIQQINRPSNVAHILQSLSAPTKSLEQQVARGQQGHPNASAVLFGQAKGAPE
THVLQQHHPPQQPQQQHPALHLQPQIMQLQQQQQQQQQQQQQPQPYPQPPSHQFPQQV
HQHQFSQQQLQFPQQPLHPQQQLHRPQQQLQPFQQQHALQQQLHQLQQQQLQHHQLAQ
LQQQQQQQHNLLQQQQQQQQLQRLQQQQQMQNQAAHLSQASQALQHQVLPQQPLQLSL
QPPPQQQQQQQLFGHDPAVEIPEESFLLGCVFAIADYPEQMSDKQLLATWKRIIQAHG
GTVDPTFTSRCTHLLCASQVSSMYTQALRERKRCVTAHWLNTVLKKKKLMPPHRALHF
PVAFPPGGKPCSQHIISVTGFVDNDRDDLKLMAYLACAKYTGYLCRSNTVLICKEPSG
LKYEKAKEWRIPCVNAQWLGDILLGNFEALRQVQYSRYTAFNMPDPFVPTPHLVLGLL
DAWRTPVKVTAELLMGVRLXPKLKPNEVANIQLSSKRARIEDLPPPTKKLTPELTPLV
LFTGFEPVQVQQYIKKLYILGGEVAECTKKCTHLIASKVTRTVKFLTAISVVKHIVTP
DWLEECFKRQTFIDEQNYILRDAEAEVLFSFSLEESLKRAHVSPFFKTKYFYITPGIC
PSLATMKAIVECAGGKVLAKQPSFRKLMEHKQNKSLSEIILISCENDLHLCREYFARG
IDVHNAEFVLTGVLTQTLDYESYKFN"
BASE COUNT 968 a 1058 c 1012 g 840 t 4 others
ORIGIN

Query Match 41.4%; Score 1483; DB 10; Length 3882;
Best Local Similarity 68.2%; Pred. No. 0;
Matches 2553; Conservative 0; Mismatches 643; Indels 546; Gaps 14;

```
Qy    175 acctctaggcacaggttattcagcttctcaaggctggaaaagcgaaggaagtttcctaca 234
          || || |  | |||||||||||||||| || || ||||| ||||| ||||||| |||||||
Db    339 ACATCGACCCCGCAGGTTATTCAGCTCCTTAAAGCTGGGAAAGCCAAGGAAGTCTCCTACA 398

Qy    235 atgcactagcctcacacataatctcagaggatggggacaatccagaggtgggagaagctc 294
          |||| || |||||||||| || ||||||||||||||||||| ||||||||||||||||| |
Db    399 ATGCCCTGGCCTCACACATCATTTCAGAGGATGGGGACAACCCAGAGGTGGGAGAAGCCC 458

Qy    295 gggaagtctttgacttacctgttgtaaagccttcttgggtgattctgtccgttcagtgtg 354
          | || |||||||||||| || |||||||||||| |||||||  |||||||| || ||||
Db    459 GTGAGGTCTTTGACTTGCCCGTTGTAA.CCTTCCTGGGTGACTCTGTCCGTGCAATGTG 518

Qy    355 qaactcttctgccagtaaatggttt'tctcc agaatcatgtcagatttttttggaatca 414
          | |||| |||||||| |||||||| |||||| || |||||||||||||||||| |||
Db    519 GCGCTCTCCTGCCAGTCAATGGTTTCTCTCCAGAGTCCTGTCAGATTTTTTTTGGACTCA 578

Qy    415 ctgcctgccttttctcagggtgttgatacaagctggagctcttttgttggagtcttccagag 474
          ||||||||||||||||
Db    579 CTGCCTGCCTTTCTCA---------------------------------------------- 594

Qy    475 ctctcccaaggagaggtagggaagggagcttgtccagcagaagttgggaagcacagagat 534
Db    595 ---------------------------------------------------------- 594

Qy    535 catctgccttcttctgacccggtattgatgcaggctgaggcctctgttgtaatgtgctgg 594
          |
Db    595 ----------------------------------------------------------G 595

Qy    595 gtgtcatctgaagacagaagtgccctgtgggcttggttacgttctatgggggagattgc 654
          |||||||||||||||||||||||||||||||||||||||||| |||||||||||| | |||||   ||||
Db    596 GTGTCATCTGAAGACAGAAGTGCCCTGTGGGCCTTGGTTACGTTCCACGGGGGCAGTTGC 655

Qy    655 cagctaaccctcaataagaaatgcacgcatttgattgttccagagccaaagggggggaaa 714
          || |||| ||| || ||||| ||||| || |||||||| |||||||||||||||||| ||||
Db    656 CAACTAAACCTTAACAAGAAGTGCACACACTTGATTGTCCCAGAGCCAAAGGGGGAGAAA 715

Qy    715 tacgaatgtgctttaaagcgagcaagtattaaaattgtgactcctgactgggttctggat 774
          |||||  |||| | ||  || |||||||||| ||||| || || |||||||| |||||
Db    716 TACGAGCGTGCTGTCAAAACAACAAGTATTAAGATTGTCACCCCGGACTGGGTCCTGGAC 775

Qy    775 tgcgtatcagagaaaaccaaaaagqacgaagcattttatcatcctcgtctgattatttat 834
          || || || ||||| |  |  ||||| ||||| ||||| ||||| || ||||| || |||
```

```
Db    776 TGTGTGTCTGAGAAGAGGAGGAAGGATGAAGCGTTTTACCATCCACGCCTGATCATCTAT 835

Qy    835 gaagaggaagaagaggaagaggaagaggaggaggaagtagaaaatgaggaacaagattct 894
          || ||||| |||||||| |||||||||          | ||| || ||||| || ||||||
Db    836 GAGGAGGAGGAAGAGGAGGAGGAAGAG---------GGAGACAACGAGGAGCAGGATTCC 886

Qy    895 cagaatgagggtagtacagatgagaagtcaagccctgccagctctcaaa..gggtctcct 954
          ||||| |||| || ||  |||||||| ||    |||||||||  || |  || |||
Db    887 CAGAACGAGGGCAGCAC---TGAGAAGTCCAGTGTCGCCAGCTCTGCAGTGGCCTCCCCT 943

Qy    955 tcaggtgaccagcagttttcacctaaatccaacactgaaaaatctaaagggaatttaatg 1014
            | ||| |||| ||  || || ||  |||   ||||    || ||||| || || |||
Db    944 ---GCTGAGCAGCCGTGCTCTCCCAAGCCCAGGGCTGAGGTGTCCAAAGGAGAGTTGATG 1000

Qy    1015 tttgatgattcttcagattcatcaccggaaaaacaggagagaaatttaaactggacccca 1074
           |||||||| ||||| || ||||| || ||||| |||||| | |  || |||||| ||||
Db    1001 TTTGATGACTCTTCTGACTCATCCCCTGAAAAGCAGGAGCGCAGCTTGAACTGGGCCCCT 1060

Qy    1075 gccgaagtcccacagttagctgcagcaaaacgcaggctgcctcagggaaaggaqcctqgg 1134
           || |||| |||||    |      |||| | ||||||||||||||||||||||| | ||| ||
Db    1061 GCTGAAGCCCCACCACTCAACACAGCCCAGCGCAGGCTGCCTCAGGGAAAAGGGCCAGGA 1120

Qy    1135 ttgattaacttgtgtgccaatgtcccacccgtcccaggtaacattttgccccctgaggtc 1194
            | |||||| |||||||||||||||||||| || |||||  ||||| |||| ||||| |
Db    1121 CTCATTAACCTGTGTGCCAATGTCCCACCTGTTCCAGGGGACATTCTGCCTCCTGACATG 1180

Qy    1195 cggggtaatttaatggctgctggacaaaacctccaaagttctgaaagatcagaaatgata 1254
           |||||||| || |||||  |||||||||||||||| | ||||||  | ||||||||  ||
Db    1181 CGGGGTAACTTGATGGCCCCTGGACAAAACCTCCAGAATTCTGAGCGGTCAGAAATACTA 1240

Qy    1255 gctacctggagtccagctgtacggacactgaggaatattactaataatgctgacattcag 1314
           | |||||||| |||||||| ||||| || || ||||| || |||||||| || |||
Db    1241 GGGACCTGGAGCCCAGCTGTCCGGACGTTGCGAAACATTACAAACAATGCTGATATCCAG 1300

Qy    1315 cagatgaaccggccatcaaatgtagcacatatcttacagactctttcagcacctacgaaa 1374
           ||||| |||||||||| ||||| |||||||||||||| | |||||||||| || |||
Db    1301 CAGATCAACCGGCCATCCAATGTTGCACATATCTTACAATCCCTTTCAGCACCCACAAAA 1360

Qy    1375 aatttagaacagcaggtgaatcacagccagcagggacatacaaatgccaatgcagtgctg 1434
           |  |||| ||||||||   |  ||||||||||||||  | |||||||| ||| ||||||
Db    1361 AGCCTAGAGCAGCAGGTGGCCCGTGGCCAGCAGGGACACCCCAATGCCAGTGCTGTGCTG 1420

Qy    1435 tttagccaagtgaaagtgactccagagacacacatgctacagcagcagcagcaggcccag 1494
           ||| |||| |  ||||  |||| |||||||||| || | |||||||| || |  |||||
Db    1421 TTTGGCCAGGCTAAAGGTGCTCCCGAGACACACGTGTTGCAGCAGCACCACCCTCCCCAG 1480

Qy    1495 cagcagcagcagcagcacccggttttacaccttcagccccagcagataatgcagctccag 1554
           ||||  |||||||||||||| |  | ||||| |||||||||  ||  ||  | |||
Db    1481 CAGCCTCAGCAGCAGCACCCAGCCCTGCACCTCCAGCCCCAGATCATGCAGCTCCAGCAG 1540

Qy    1555 cagcagcagcagcagca---------------------------------------------- 1571
           ||||||||||||||||
Db    1541 CAGCAGCAGCAGCAGCAACAGCAGCAGCAGCAGCCCCAGCCCTACCCCCAGCCTCCATCA 1600

Qy    1572 ---------------------------------------------------------- 1571
Db    1601 CATCAGTTCCCGCAGCAAGTCCATCAGCATCAGTTCTCTCAGCAGCAGCTCCAGTTTCCA 1660

Qy    1572 ----------------------------------------------------------gatc 1575
                                                                       |
Db    1661 CAGCAACCATTACACCCTCAGCAGCAGTTGCATCGTCCTCAACAGCAGCTGCAGCCATTT 1720

Qy    1576 tctcagcaaccttacccccagcagccgccgcatccattttcacagcagcagcagcagcaa 1635
              |||| | |  ||  |||||| ||  |||| ||    |||||||||  |||||
Db    1721 CAGCAGCAGCATGCCCTGCAGCAGCAGCTTCATCAGTTGCAGCAGCAGCAGCTCCAGCAC 1780

Qy    1636 cagcag---------------------------------------------------- 1641
           || |||
Db    1781 CACCAGTTAGCACAGCTCCAGCAACAGCAGCAACAGCAGCACAACCTTCTCCAGCAGCAG 1840

Qy    1642 ---------------------------------------------------------- 1641
Db    1841 CAACAACACCAGCAGCTCCAACGCTTGCAGCAGCAGCAGCAGATGCAGAATCAAGCAGCA 1900
```

```
Qy   1642 ------------------------------------------------------------ 1641

Db   1901 CACTTGAGCCAAGCGTCCCAGGCACTACAGCACCAGGTTCTACCCCAGCAGCCCCTGCAG 1960

Qy   1642 ---------------ccaccaccatcgcctcagcagcatcagctttttggacatgatcca 1686
                          ||||| |     | |||||||| ||||||||||||||||||| |||
Db   1961 CTGTCCCTGCAGCCCCCACCTCAGCAACAGCAGCAGCAGCAGCTTTTTGGACATGACCCA 2020

Qy   1687 gcagtggagattccagaagaaggcttcttattgggatgtgtgtttgcaattgcggattat 1746
          |||||||||||||||||||||| |||||||  | ||||||||||||||||||||| || |||
Db   2021 GCAGTGGAGATTCCAGAAGAAAGCTTCTTGCTAGGATGTGTGTTTGCAATTGCAGACTAT 2080

Qy   1747 ccagaagcagatgtctgataagcaactgctggccacctggaaaaggataatccaggcacat 1806
          || |||||||||||| |||||||| ||||||||||||||||||||||||||||||||||||
Db   2081 CCCGAGCAGATGTCGGATAAGCACCTGCTGGCCACCTGGAAAAGGATAATCCAGGCACAT 2140

Qy   1807 ggcggcactgttgaccccacctcacgagtcgatgcacgcaccttctctgtgagagtcaa 1866
          || |||||||| |||||| || |||| || ||||| || ||||| |||||  | |||||||
Db   2141 GGTGGCACTGTGGACCCCACTTTCACAAGCCGATGTACACACCTCCTCTGCGCAAGTCAA 2200

Qy   1867 gtcagcagcgcgtatgcacaggcaataagagaaagaaagagatgtgttactgcacactgg 1926
          ||||| ||   |||  |||||||| || | || |||||| | ||||| ||||||||||||
Db   2201 GTCAGTAGTATGTACACACAGGCACTACGGGAGAGAAAGCGGTGTGTCACTGCACACTGG 2260

Qy   1927 ttaaacacagtcttaaaaaaaaaaaaaatggtaccgccgcaccgagcccttcacttccca 1986
           | ||||||||  | || || || ||  || | || || || |.|||||| || |||||
Db   2261 CTCAACACAGTGCTGAAGAAGAAGAAGCTGATGCCCCCTCATCGAGCCCTGCATTTCCCG 2320

Qy   1987 gtggccttcccaccaggaggaaagccatgttcacagcatattatttctgtgactggattt 2046
          ||||| |||||||| || || || || ||||||||||||||||||||| ||||||||||||
Db   2321 GTGGCGTTCCCACCTGGGGGCAAACCCTGTTCACAGCATATTATCTCTGTGACTGGATTT 2380

Qy   2047 gttgatagtgacagagatgacctaaaattaatggcttatttggcaggtgccaaatatacg 2106
          ||||| | |||||||||||||||| || |||||||| ||  |||||||||||||| || ||
Db   2381 GTTGACAATGACAGAGATGACCTGAAGTTAATGGCATACCTGGCAGGTGCCAAGTACACA 2440

Qy   2107 ggttatctatgccgcagcaacacagtcctcatctgtaaagaaccaactggtttaaagtat 2166
          || || |||| |||||||||||||||||||||||||||||||||  || || ||||||
Db   2441 GGGTACCTGTGCCGTAGCAACACAGTCCTCATCTGTAAAGAACCAAGCGGCTTGAAGTAT 2500

Qy   2167 gaaaaagccaaagagtggaggatacccctgtgtcaacgcccagtggcttggcgacattctt 2226
          ||||| |||||||||||||||||| || |||| || || |||||||| || |||||||||
Db   2501 GAAAAGGCCAAAGAGTGGAGGATCCCATGTGTGAATGCACAGTGGCTGGGAGACATTCTG 2560

Qy   2227 ctgggaaactttgaggcactgaggcagattcagtatagtcgctacacggcattcagtctg 2286
          ||||| |||||||||||| ||| ||||  | ||||| || || ||||| || ||||  ||
Db   2561 CTGGGGAACTTTGAGGCCCTGCGGCAAGTGCAGTACAGCCGGTACACCGCCTTCAACATG 2620

Qy   2287 caggatccatttgcccctacccagcatttagttttaaatcttttagatgcttggagagtt 2346
          | ||| || || |  || || |  || || ||   |||| ||||||| |||||
Db   2621 CCGGACCCCTTCGTTCCCACTCCACACTTGGTCTTGGGTCTTCTAGATGCATGGAGGACC 2680

Qy   2347 cccttaaaagtgtctgcagagttgttgatgagtataagactacctcccaaactgaaacag 2406
          ||  | || ||| |||||||| |||||||| || ||||| ||| ||| || |||||||| |
Db   2681 CCAGTGAAGGTGACTGCAGAGCTGTTGATGGGTGTAAGATTACNTCCTAAGCTGAAACCG 2740

Qy   2407 aatgaagtagctaatgtccagccttcttcccaaagagccagaattgaagacgtaccacct 2466
          ||||| || |||||  |||||| ||||||  |||||||||||||||||||| | ||||||
Db   2741 AATGAGGTCGCTAACATCCAGCTTTCTTCTAAAAGAGCCAGAATTGAAGACTTGCCACCT 2800

Qy   2467 cccactaaaaagctaactccagaattgacccctttgtgcttttcactggattcgagcct 2526
          |||||||| ||||| ||||||||  |||| |||| |||||  |||||||| || ||||||
Db   2801 CCCACTAAGAAGCTGACTCCAGAGCTGACGCCTTTAGTGCTCTTCACTGGGTTTGAGCCT 2860

Qy   2527 gtccaggttcaacagtatattaagaagctctacattcttggtggagaggttgcggagtct 2586
          || || ||||| ||||| |||||||||||  |||||||||||||| || ||||| ||||
Db   2861 GTTCAAGTTCAGCAGTACATTAAGAAGCTGTACATTCTTGGTGGGGAAGTTGCTGAGTGC 2920

Qy   2587 gcacagaagtgcacacacctcattgccagcaaagtgactcgcaccgtgaagttcctgacg 2646
          |  ||||||| || |||||||||||||||| ||||| ||||| ||||| ||| ||||||
Db   2921 ACGAAGAAGTGTACCCACCTCATTGCCAGCAAGGTGACACGCACAGTGAAATTCTTGACG 2980
```

```
Qy   2647 gcgatttctgtcgtgaagcacatagtgacgccagagtggctggaagaatgcttcaggtgt 2706
          || || |||||| |||||||| |||||| || || |||||||| || |||||| | |
Db   2981 GCCATCTCTGTAGTAAAGCACATTGTGACTCCGGACTGGCTGGAGGAGTGCTTCAAGCGC 3040

Qy   2707 cagaagttcattgatgagcagaactacattctccgagatgctgaggcagaagtacttttc 2766
          |||| || || |||||||||||||||||| || |||||||||||||||||||| || |||
Db   3041 CAGACATTTATCGATGAGCAGAACTACATCCTTCGAGATGCTGAGGCAGAAGTGCTCTTC 3100

Qy   2767 tctttcagcttggaagaatccttaaaacgggcacacgtttctccactctttaaggcaaaa 2826
          ||||||||  |||||||| |||||||| || || || || |||||  |||||||| ||||
Db   3101 TCTTTCAGTCTGGAAGAGTCCTTAAAGCGAGCCCATGTCTCTCCCTTCTTTAAGACAAAG 3160

Qy   2827 tatttttacatcacacctggaatctgcccaagtctttccactatgaaggcaatcgtagaq 2886
          ||||| || |||| |||||||||||| ||| | ||||||||||||| ||||||||||||
Db   3161 TATTTCTATATCACCCCTGGCATCTGCCCAAGCCTTGCAACTATGAAGGCAATTGTAGAG 3220

Qy   2887 tgtgcaggaggaaaggtgttatccaagcagccatctttccggaagctcatggagcacaag 2946
          |||||||| |||||||| || |||||||||| || |||||||||||||||||||||||||
Db   3221 TGTGCAGGGGGAAAGGTGCTAGCCAAGCAGCCGTCCTTCCGGAAGCTCATGGAGCACAAG 3280

Qy   2947 cagaactcgagtttgtcggaaataatttttaatatcctgtgaaaatgaccttcatttatgc 3006
          |||||   |||| | || || || |||||||||||||||||| || ||||| || |||||
Db   3281 CAGAATAAGAGTCTTTCAGAGATCATTTTAATATCCTGTGAGAACGACCTGCACTTATGT 3340

Qy   3007 cgagaatattttgccagaggcatagatgttcacaatgcagagttcgttctgactggagtg 3066
          || || || |||||||| || |||||||||||||||||| ||||| || ||||||||||||
Db   3341 CGCGAGTACTTTGCCAGGGGAATAGATGTTCACAATGCGGAGTTTGTCCTGACTGGAGTA 3400

Qy   3067 ctcactcaaacgctggactatgaatcatataagtttaactgatggcgtctaggctgccgt 3126
          ||||| || || |||||||||||||||||||||||||||||||||||| |||  ||| |
Db   3401 CTCACACAGACACTGGACTATGAATCATATAAATTTAACTGATGGCCTCT----TGCTGC 3456

Qy   3127 gcatgtcgactcctgcggtgcggggctggctgtctggctggcgaggagctgctgcgcttc 3186
             |  |         || ||       |   |  ||  | | |  | |  ||
Db   3457 TGCTAGCCTCGAGCCTGGAGCACAAGGGTAGCTGCTGGCAGCATGAACCAGGACTCCCTC 3516

Qy   3187 cttcacatgctcttgttttccagctgctttcctgggggatcagactgtgaagcaggaaga 3246
          |       | |||||||||||||||         ||| ||     |    | ||||||||
Db   3517 CCCGCAGAGGGCTTGTTTTCCAGCTCTGCCTTAGGGAGAGACCTCG----ACCAGGAAGA 3572

Qy   3247 cagatataataaatatactgcatctttttaagatgtgcaatttttattctgaggaaacata 3306
          ||||| |||||     |||||||| | |||||||||  || |||||  |||||||||||
Db   3573 CAGATGTAATACTG--ACTGCATCGTGTTAAGATGTGTGATCTTATTNT-AGGAAACATA 3629

Qy   3307 aattatgttttgtattatatgactttaagagcccacattaggttttatgattcatttgcc 3366
          |||||||||||||||||||| | ||||||||||||||| ||||| ||||| |||||||
Db   3630 AATTATGTTTTGTATTATATGCC--TAAGAGCCCACATTAAGTTTTCTGATTAATTTGCC 3687

Qy   3367 aggtttttaaatgttttcacaaaactgttacgggacttcaactagaaataaaatggtgta 3426
          ||||||||||    |||    |||   | | | |||| || |||||||| |
Db   3688 AGGTTTTTAATGTGTTTTCAGAAAGCTGTGTGAAATTTCATTTACAAATAAAAATGATGT 3747

Qy   3427 aataaagaccttgctatctctaaattatggatgttaaagatttgaaatgttttgtacttt 3486
          ||  ||||| |||| | | |||||||||| |||||||||||||||||||||||||||||
Db   3748 AAATAAGACTTTGC--TGTNTAAATTATGGCTGTTAAAGATATGAAATGTTTTGTACTTT 3805

Qy   3487 gattatttttatttcttatactctgttttctttttatattgatatcttgcccacattttaa 3546
          |||||||||||||||||| ||    |||||||||||||||||| ||||| ||||  ||||
Db   3806 GATTATTTTTATTTCTTACAC----TTTTCTTTTATATTGATAACTTGCTCACA-GTTAA 3860

Qy   3547 ataaaatgtacttttgaacttag 3568
          |||||||||||||||||||||||
Db   3861 ATAAATGTACTTTTGAACTTAG 3882
```

RESULT 4
AF172855
LOCUS AF172855 3771 bp mRNA linear VRT 22-MAY-2001
DEFINITION Xenopus laevis Swift (K14) mRNA, complete cds.
ACCESSION AF172855
VERSION AF172855.1 GI:14164560
KEYWORDS .
SOURCE African clawed frog.

```
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 3771)
  AUTHORS   Shimizu,K., Bourillot,P.Y., Nielsen,S.J., Zorn,A.M. and Gurdon,J.B.
  TITLE     Swift Is a Novel BRCT Domain Coactivator of Smad2 in Transforming
            Growth Factor beta Signaling
  JOURNAL   Mol. Cell. Biol. 21 (12), 3901-3912 (2001)
   PUBMED   11359898
REFERENCE   2  (bases 1 to 3771)
  AUTHORS   Shimizu,K. and Gurdon,J.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JUL-1999) Wellcome/CRC Institute, Tennis Court Road,
            Cambridge CB2 1QR, UK
FEATURES             Location/Qualifiers
     source          1. .3771
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /tissue_type="embryo"
     gene            1. .3771
                     /gene="K14"
     CDS             1. .3771
                     /gene="K14"
                     /note="nucleoprotein"
                     /codon_start=1
                     /product="Swift"
                     /protein_id="AAK55123.1"
                     /db_xref="GI:14164561"
                     /translation="MAEDEVKVPEDMFKDVKFYPVGDLDPKVVQMLKAGKAKEVSYNA
                     LATHIIAEDGDNPEVGEAREVFDLPVVKPSWVILSVRCGVRLPENGFSPESGQIFFGV
                     TACLSQVSPDDRNSLWALTTFYGGDCQLSLNKKCTHLIVPEPKGNKYEYAFQRGSIKI
                     VTPDWVLDSVSEKTKKDEALYHPRLIVYEEEEEEEEEEAENEEQESQNGSDEDKYSSK
                     SSPVTSRDGSPVHNRESSPKRSSSDKIKSELMFDDSSDSSPEKQERNLNWTPAEVPQM
                     STAKRQLHQGAPGSERPDMMAFRSPAVRTLRNITNSADVQQVNRPSNVAHILQTLSAS
                     TKSLEQRVNHPQQGHPNAVLFGQVKPLTSEAQHLIQQSHQPHHQQQQQQQNHPIVQLQ
                     PQQLLQLQQQQQQITQQVFPQHQFPQVNQQHPFTQLQFPQQQLHPQQQLHRPQQQTIQ
                     HFQQQHALQQQLHQIQQQHLQPKPQTLQQNMQQQNLQQPNLQQILQHQQTIQQTPSQQ
                     QALQPAIQQQQMLQPNIQQQQTLQSNLQQQTLQPIIQQQQGLQQQNLQQSLQQIQQQM
                     QHLTPQQKQQIQQQQQQMLQKQQLQSQSLQQQQMQTHVLQQQQIQTQALQQQQQIQNQ
                     ALQQPQQNQVQTHALQQHRLQSQTLQQQHHVQGQTVQQQTHQLQTQTLQQQHQIQTLQ
                     IPHQIPAPNQQHQIPPQMLQQQTLQLQQQMQPQIQQPQMQSGVQQQSSLQPQQMQQHK
                     HNLQQVQHQLQQLQQQRMQQRQMALPQQIANQQPPQPHQLHGHDPSVEIPEDFFLLGC
                     VFAIADYPEQMPDKQLLATWKRIIQTHGGTVDAALSSRCTHLLCESQVSSMYAQALKE
                     RKRCITAHWLNAVLKKKKMVPPYRALHFPVAFPPGGKPCSQHIISVTGFVDSDRDDLK
                     LMGYLAGAKYTGYLCRSNTVLICKEPSGLKYEKAKEWRIPCVNALWLCDILLGNLEAL
                     RQIQHSRYTVFNLQDPLAPSPHLVTDDLLDAWRMPLKVSSEVLMSIRMPLKPKQNEPA
                     VQPKRPRIEDIPPPTKKLSPDQTPHVIFTGFDPLQVQQYIKKLYILGGEVADTAQKCT
                     HLVANKVTRTVKFLTAISVAKHIVTPEWLDESFKSQKFAEEQNYILRDAEAEVLFCFS
                     LEESLKKAHVNPLFKGKYFYITPGICPSLSTMKAIVECAGGKILTKQPSFRKIMEHKQ
                     NKRLAEIILISCENDLHLCREYFAGSVDVHNAEFVLTGVLTQALDYESYKFT"
BASE COUNT     1189 a    931 c    792 g    858 t      1 others
ORIGIN
```

```
  Query Match           24.4%;  Score 874;  DB 5;  Length 3771;
  Best Local Similarity 71.8%;  Pred. No. 1.6e-180;
  Matches 1189;  Conservative    0;  Mismatches  451;  Indels   15;  Gaps    3;

Qy   1460 agacacacatgctacagcagcagcagcaggcccagcagcagcagcagcagcacccggttt 1519
          ||  ||| |||  |  |||   ||||  | |||  ||||          || |   |
Db   2126 AGCAACAAATGCAGCCNCAGATTCAGCAACCGCAGATGCAGTCAGGAGTTCAGCAACAGT 2185

Qy   1520 tacaccttcagccccagcagatаatgcagctccagcagcagcagcagcagcagatctctc 1579
             || |||| ||||||||    || |   | ||  | | |||||||         ||
Db   2186 CCTCACTCCAGCCACAGCAGATGCAACAACATAAACACAATCTGCAGCAGGTTCAGCATC 2245

Qy   1580 agcaaccttacccccagcagccgccgcatccatttтcacagcagcagcagcagcaacagc 1639
          | |  |   | |  || |||| || |   |       |||   |      | ||||||
Db   2246 AACTGCAACAACTTCAACAGCAGCGGATGCAGCAGCGACAAATGGCATTACCCCAACAGA 2305

Qy   1640 agcc---accaccatcgcctcagcagcatcagctttttggacatgatccagcagtggaga 1696
               |    |  | | | || |||| |||||||||    || ||||||||  | |||||||

Db   2306 TTGCTAATCAGCAACCACCACAGCCGCATCAGCTGCACGGTCATGATCCTTCTGTGGAGA 2365

Qy   1697 ttccagaagaaggcttcttattgggatgtgtgtttgcaattgcggattatccagagcaga 1756
          ||||  |||||   ||| || ||||||||||||||||||||||||| ||||||||||| || |
Db   2366 TTCCTGAAGATTTCTTTTTGTTGGGATGTGTGTTTGCAATTGCTGATTATCCAGAACAAA 2425

Qy   1757 tgtctgataagcaactgctggccacctggaaaaggataatccaggcacatggcggcactg 1816
          || | ||||| |||| || || |||||||||| |||||| ||| | |||| || || | |
Db   2426 TGCCCGATAAACAACTACTTGCAACCTGGAAACGGATAATACAGACCCATGGAGGTACGG 2485

Qy   1817 ttgaccccaccttcacgagtcgatgcacgcaccttctctgtgagagtcaagtcagcagcg 1876
          | ||| |  |  |  | ||  | |||| || |||| |||| || ||||||| |||
Db   2486 TGGACGCAGCACTTTCAAGCAGGTGCACTCATCTTCTGTGTGAAAGCCAAGTCAGTAGCA 2545

Qy   1877 cgtatgcacaggcaataagagaaagaaagagatgtgttactgcacactggttaaacacag 1936
           |||||| |||||| | | ||| || ||| |||| | |||| || ||| | ||  |||
Db   2546 TGTATGCTCAGGCATTGAAAGAGAGGAAGCGATGTATCACTGCTCATTGGCTGAATGCAG 2605

Qy   1937 tcttaaaaaaaaaaaaaatggtaccgccgcaccgagcccttcacttcccagtggccttcc 1996
          | || |||| || |||||||| |||||| |||||| |||| |||| || || | |
Db   2606 TATTGAAAAAGAAGAAAATGGTGCCGCCGTACCGAGCTCTTCATTTCCCTGTAGCATTTC 2665

Qy   1997 caccaggaggaaagccatgttcacagcatattatttctgtgactggatttgttgataqtg 2056
          |||||||||| |||| || || |||||||| |||| |||| || |||||||| ||||
Db   2666 CACCAGGAGGGAAGCCCTGCTCTCAGCATATCATTTCAGTGACCGGGTTTGTTGACAGTG 2725

Qy   2057 acagagatgacctaaaattaatggcttatttggcaggtgccaaatatacgggttatctat 2116
          |||||||||| || ||| ||||||  || || |||| |||| || || || || || | |
Db   2726 ACAGAGATGATCTGAAACTAATGGGGTACTTAGCAGGCGCCAAGTACACTGGATACCTGT 2785

Qy   2117 gccgcagcaacacagtcctcatctgtaaagaaccaactggtttaaagtatgaaaaagcca 2176
          | |||||||||| || |||| |||||||| || |  || || |||| || |||||| |
Db   2786 GTCGCAGCAACACTGTTCTCATTTGTAAAGAGCCGAGCGGCTTGAAGTACGAGAAAGCTA 2845

Qy   2177 aagagtggaggataccctgtgtcaacgcccagtggcttggcgacattcttctgggaaact 2236
          |||||||||||||||| |||| || |||| |||||||| | || |||||||| ||||||
Db   2846 AAGAGTGGAGGATACCGTGTGTAAATGCCCTGTGGCTTTGTGATATTCTTCTTGGAAACC 2905

Qy   2237 ttgaggcactgaggcagattcagtatagtcgctacacggcattcagtctgcaggatccat 2296
          |||||||| |  | |||||||| | || || || || | ||| | ||| |||| ||
Db   2906 TTGAGGCATTACGTCAGATTCAGCACAGCCGATATACAGTATTTAATCTTCAGGACCCTC 2965

Qy   2297 ttgcccctacccagcatttagtttta---aatcttttagatgcttggagagttcccttaa 2353
          |||| ||  | |  ||| |  |||| |||||| |||||| | || || |
Db   2966 TTGCACCATCACCTCATCTGGTTACAGATGATCTCCTAGATGCCTGGAGAATGCCATTGA 3025

Qy   2354 aagtgtctgcagagttgttgatgagtataagactacctcccaaactgaaacagaatgaag 2413
          |||||||  | ||  | |||||||||| || | |||  || | || || ||||||
Db   3026 AAGTGTCCTCGGAAGTTCTGATGAGTATTCGAATGCCTCTAAAGCCGAAGCAAAATGAGC 3085

Qy   2414 tagctaatgtccagccttcttcccaaagagccagaattgaagacgtaccacctcccacta 2473
            |||  |   |          | || |  | || |||||||  |||| |||||  ||||
Db   3086 CTGCTGTTCAAC---------CTAAACGCCCAAGGATTGAAGATATACCGCCTCCAACTA 3136

Qy   2474 aaaagctaactccagaattgacccctttgtgcttttcactggattcgagcctgtccagg 2533
          |||||||  | || ||  ||| ||| |||  | ||||||||||| || ||  | ||||
Db   3137 AAAAGCTGTCACCTGATCAGACACCTCATGTTATATTCACTGGATTTGATCCACTGCAGG 3196

Qy   2534 ttcaacagtatattaagaagctctacattcttggtggagaggttgcggagtctgcacaga 2593
          | || || ||||| |||| || || |||||||| |||| || || ||  | |||||||
Db   3197 TCCAGCAATATATCAAGAAACTGTATATTCTTGGCGGAGAAGTAGCCGATACGGCACAGA 3256

Qy   2594 agtgcacacacctcattgccagcaaagtgactcgcaccgtgaagttcctgacggcgattt 2653
          | |||| ||| |  |||||| |||||||| || || || |||||| |||| || ||||
Db   3257 AATGCACTCACTTGGTTGCCAACAAAGTGACCCGGACTGTAAAGTTCTTGACTGCAATTT 3316

Qy   2654 ctgtcgtgaagcacatagtgacgccagagtggctggaagaatgcttcaggtgtcagaagt 2713
          | || |  |||||||| || || || |||||||| || ||  ||||||  |  || ||||
Db   3317 CAGTGGCAAAGCACATTGTTACCCCCGAGTGGCTAGATGAGAGCTTCAAATCCCAAAAGT 3376

Qy   2714 tcattgatgagcagaactacattctccgagatgctgaggcagaagtacttttctctttca 2773
          |   ||| || |||||  | || |||||||| || |||||  || |||| ||||
Db   3377 TTGCTGAGGAACAGAATTACATATTGCGGGATGCTGAAGCGGAAGTGCTGTTCTGCTTCA 3436
```

```
Qy   2774 gcttggaagaatcccttaaaacgggcacacgtttctccactctttaaggcaaaatattttt 2833
          | || ||||||||  |||||   || |||||   ||| ||||| ||||  || ||||||||
Db   3437 GTTTAGAAGAATCGCTAAAAAAAGCGCACGTGAATCCGCTCTTCAAGGGGAAGTATTTTT 3496

Qy   2834 acatcacacctggaatctgcccaagtctttccactatgaaggcaatcgtagagtgtgcag 2893
          ||||||| |||||||| |||||||||||| || || ||||| || || ||||| ||||| |
Db   3497 ACATCACCCCTGGAATTTGCCCAAGTCTCTCGACAATGAAAGCCATTGTAGAATGTGCTG 3556

Qy   2894 gaggaaaggtgttatccaagcagccatctttccggaagctcatggagcacaagcagaact 2953
          | || ||  |  || | |||||||| ||||| |||||  | ||||| || ||||||||
Db   3557 GTGGGAAAATACTAACGAAGCAGCCGTCTTTTCGGAAAATTATGGAACATAAGCAGAATA 3616

Qy   2954 cgagtttgtcggaaataatttttaatatcctgtgaaaatgaccttcatttatgccgagaat 3013
          | ||  | || || |||||||||||||| |||||||| |||||| | ||| ||||||
Db   3617 AACGATTAGCTGACATCATTTTAATATCCTGCGAAAATGATCTTCATCTGTGCAGAGAAT 3676

Qy   3014 attttgccagaggcatagatgttcacaatgcagagttcgttctgactggagtgctcactc 3073
          | |||||  |  || | ||||| |||||||||| || ||| |||| ||||| ||||| |
Db   3677 ACTTTGCTGGCAGCGTGGATGTGCACAATGCAGAATTTGTTTTGACGGGAGTTCTCACCC 3736

Qy   3074 aaacgctggactatgaatcatataagtttaactga 3108
          |  | ||||||||||||||||||||||||||  |||
Db   3737 AGGCACTGGACTATGAATCATATAAGTTTACTTGA 3771


RESULT   5
AC073077
LOCUS       AC073077               149592 bp    DNA     linear   HTG 17-NOV-2000
DEFINITION  Homo sapiens chromosome 7 clone RP11-269M19, WORKING DRAFT
            SEQUENCE, 10 unordered pieces.
ACCESSION   AC073077
VERSION     AC073077.5  GI:11192192
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 149592)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 149592)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JUN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Nov 17, 2000 this sequence version replaced gi:11054999.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0269M19
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 18%
            Sequencing vector: plasmid; 64%
            Chemistry: Dye-primer ET; 18% of reads
            Chemistry: Dye-terminator Big Dye; 64% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 142511 bases at least Q40
            Consensus quality: 144983 bases at least Q30
            Consensus quality: 146462 bases at least Q20
            Insert size: 167000; agarose-fp
            Insert size: 148692; sum-of-contigs
            Quality coverage: 5.83 in Q20 bases; agarose-fp
            Quality coverage: 5.97 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     4209: contig of 4209 bp in length
            *    4210     4309: gap of unknown length
            *    4310    25449: contig of 21140 bp in length
            *   25450    25549: gap of unknown length
            *   25550    42417: contig of 16868 bp in length
            *   42418    42517: gap of unknown length
            *   42518    60534: contig of 18017 bp in length
            *   60535    60634: gap of unknown length
            *   60635    97012: contig of 36378 bp in length
            *   97013    97112: gap of unknown length
            *   97113   139971: contig of 42859 bp in length
            *  139972   140071: gap of unknown length
            *  140072   141186: contig of 1115 bp in length
            *  141187   141286: gap of unknown length
            *  141287   143081: contig of 1795 bp in length
            *  143082   143181: gap of unknown length
            *  143182   145560: contig of 2379 bp in length
            *  145561   145660: gap of unknown length
            *  145661   149592: contig of 3932 bp in length.
FEATURES             Location/Qualifiers
     source          1. .149592
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /clone="RP11-269M19"
     misc_feature    1. .4209
                     /note="assembly_name:Contig10
                     clone_end:T7
                     vector_side:left"
     misc_feature    4310. .25449
                     /note="assembly_name:Contig11"
     misc_feature    25550. .42417
                     /note="assembly_name:Contig12"
     misc_feature    42518. .60534
                     /note="assembly_name:Contig13
                     clone_end:SP6
                     vector_side:left"
     misc_feature    60635. .97012
                     /note="assembly_name:Contig14"
     misc_feature    97113. .139971
                     /note="assembly_name:Contig15"
     misc_feature    140072. .141186
                     /note="assembly_name:Contig6"
     misc_feature    141287. .143081
                     /note="assembly_name:Contig7"
     misc_feature    143182. .145560
                     /note="assembly_name:Contig8"
     misc_feature    145661. .149592
                     /note="assembly_name:Contig9"
BASE COUNT    40421 a  34230 c  33544 g  40490 t    907 others
ORIGIN


  Query Match           17.8%;  Score 636.6;  DB 2;  Length 149592;
  Best Local Similarity   98.6%;  Pred. No. 2.5e-128;
  Matches  642;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy     708 ggggaaatacgaatgtgctttaaagcgagcaagtattaaaattgtgactcctgactgggt 767
           || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   57998 GGAGAAATACGAATGTGCTTTAAAGCGAGCAAGTATTAAAATTGTGACTCCTGACTGGGT 58057

Qy     768 tctggattgcgtatcagagaaaaccaaaaaggacgaagcatttttatcatcctcgtctgat 827
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   58058 TCTGGATTGCGTATCAGAGAAAACCAAAAAGGACGAAGCATTTTATCATCCTCGTCTGAT 58117

Qy     828 tatttatgaagaggaagaagaggaagaggaagaggaggaggaagtagaaaatgaggaaca 887
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   58118 TATTTATGAAGAGGAAGAAGAGGAAGAGGAAGAGGAGGAGGAAGTAGAAAATGAGGAACA 58177
```

```
Qy    888 agattctcagaatgaggqtagtacagatgagaagtcaagccctgccagctctcaagaagg 947
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  58178 AGATTCTCAGAATGAGGGTAGTACAGATGAGAAGTCAAGCCCTGCCAGCTCTCAAGAAGG 58237

Qy    948 gtctccttcaggtgaccagcagttttcacctaaatccaacactgaaaaatctaaagggga 1007
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  58238 GTCTCCTTCAGGTGACCAGCAGTTTTCACCTAAATCCAACACTGAAAAATCTAAAGGGGA 58297

Qy   1008 attaatgtttgatgattcttcagattcatcaccggaaaaacaggagagaaatttaaactg 1067
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  58298 ATTAATGTTTGATGATTCTTCAGATTCATCACCGGAAAAACAGGAGAGAAATTTAAACTG 58357

Qy   1068 gaccccggccgaagtcccacagttagctgcagcaaaacgcaggctgcctcagggaaagga 1127
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  58358 GACCCCGGCCGAAGTCCCACAGTTAGCTGCAGCAAAACGCAGGCTGCCTCAGGGAAAGGA 58417

Qy   1128 gcctgggttgattaacttgtgtgccaatgtcccacccgtcccaggtaacattttgccccc 1187
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  58418 GCCTGGGTTGATTAACTTGTGTGCCAATGTCCCACCCGTCCCAGGTAACATTTTGCCCCC 58477

Qy   1188 tgaggtccggggtaatttaatggctgctggacaaaacctccaaagttctgaaagatcaga 1247
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  58478 TGAGGTCCGGGGTAATTTAATGGCTGCTGGACAAAACCTCCAAAGTTCTGAAAGATCAGA 58537

Qy   1248 aatgatagctacctggagtccagctgtacggacactgaggaatattactaataatgctga 1307
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  58538 AATGATAGCTACCTGGAGTCCAGCTGTACCGACACTGAGGAATATTACTAATAATGCTGA 58597

Qy   1308 cattcagcagatgaaccggccatcaaatgtagcacatatcttacagactct 1358
          ||||||||||||||||||||||||||||||||||||| |    ||  || |
Db  58598 CATTCAGCAGATGAACCGGCCATCAAATGTAGCACATGTAAGTCACTCTTT 58648
```

RESULT 6
AC093726/c
LOCUS AC093726 159041 bp DNA linear PRI 08-FEB-2002
DEFINITION Homo sapiens BAC clone RP11-5C23 from 7, complete sequence.
ACCESSION AC093726 AC013583
VERSION AC093726.3 GI:18497181
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 159041)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE 2 (bases 1 to 159041)
AUTHORS Tomlinson,C., Cotton,M. and Phillips,A.
TITLE The sequence of Homo sapiens BAC clone RP11-5C23
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 159041)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (10-SEP-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 159041)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (05-FEB-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 159041)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (08-FEB-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Feb 5, 2002 this sequence version replaced gi:15624904.
---------- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
---------- Summary Statistics
Center project name: H_NH0005C23
Drafting Center: WIBR
----------.

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see http://www.nhgri.nih.gov/DIR/GTB/CHR7 , send mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu)
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-476H24, 6 bp overlap; the clone sequenced to the right is RP11-279O22, 2000 bp overlap. Actual start of this clone is at base position 1 of RP11-5C23; actual end is at base position 23039 of RP11-279O22.

Data from AC074285 and AC073077 was used to finish this clone, AC093726. Polymorphisms have been identified between AC07428 and AC093726.

The clone RP11-5C23 contains a 112 bp insertion not sequenced in the right neighboring clone, RP11-279O22. This insertion sequence is being submitted as H_NH0005C23_F1.

The sequence of AC013583 has been incorporated into AC093726.

FEATURES	Location/Qualifiers
source	1. .159041
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="7"
	/map="7"
	/clone="RP11-5C23"
	/clone_lib="RPCI-11"
repeat_region	1. .1735
	/rpt_family="ERV1"
repeat_region	1694. .1731
	/rpt_family="(GGA)n"
repeat_region	1879. .2117
	/rpt_family="ERV1"
repeat_region	2531. .2719

```
                     /rpt_family="ERV1"
     repeat_region   2748. .3093
                     /rpt_family-"ERV1"
     repeat_region   3097. .3405
                     /rpt_family-"ERV1"
     repeat_region   3380. .3489
                     /rpt_family-"ERV1"
     repeat_region   3504. .3814
                     /rpt_family-"Alu"
     misc_feature    3994. .4270
                     /note-"similar to EST AA078489 (NID:g1837963)"
     repeat_region   4427. .4555
                     /rpt_family="Alu"
     misc_feature    4683. .5105
                     /note-"similar to EST BF212893 (NID:g11106479)"
     misc_feature    4683. .5182
                     /note="similar to EST BE748183 (NID:g10162175)"
     misc_feature    5048. .5406
                     /note="similar to EST BI034826 (NID:g14441452)"
     repeat_region   5206. .5410
                     /rpt_family-"MER1_type"
     repeat_region   5449. .5927
                     /rpt_family-"L1"
     repeat_region   5944. .5989
                     /rpt_family="(TAGA)n"
     repeat_region   6053. .6755
                     /rpt_family-"L1"
     repeat_region   6752. .6784
                     /rpt_family-"(T)n"
     repeat_region   6756. .7064
                     /rpt_family="Alu"
     repeat_region   7065. .7607
                     /rpt_family-"L1"
     repeat_region   7808. .8174
                     /rpt_family-"MaLR"
     repeat_region   8181. .9662
                     /rpt_family="MaLR"
     misc_feature    9470. .9608
                     /note="similar to EST AW855220 (NID:g7950913)"
     misc_feature    9664. .9915
                     /note="similar to EST AW855220 (NID:g7950913)"
     repeat_region   9815. .10188
                     /rpt_family="MaLR"
     repeat_region   10327. .10565
                     /rpt_family-"L1"
     repeat_region   10566. .10650
                     /rpt_family-"L1"
     repeat_region   10651. .11124
                     /rpt_family="L1"
     repeat_region   11125. .11667
                     /rpt_family-"ERVK"
     repeat_region   11668. .12906
                     /rpt_family-"L1"
     repeat_region   12921. .12945
                     /rpt_family="AT_rich"
     repeat_region   13343. .13613
                     /rpt_family-"L1"
     repeat_region   14320. .14447
                     /rpt_family-"MaLR"
     repeat_region   14451. .14488
                     /rpt_family="MaLR"
     repeat_region   14573. .14728
                     /rpt_family-"MaLR"
     repeat_region   14729. .15068
                     /rpt_family-"MER1_type"
     repeat_region   15069. .15109
                     /rpt_family="MaLR"
     repeat_region   15467. .15633
                     /rpt_family-"Alu"
     repeat_region   15611. .15633
                     /rpt_family-"(A)n"
     repeat_region   16642. .17000
                     /rpt_family="MER1_type"
     misc_feature    18037. .18430
                     /note="similar to EST AW958088 (NID:g8147771)"
     misc_feature    18039. .18514
                     /note="similar to EST BF939756 (NID:g12356999)"
     misc_feature    18040. .18557
                     /note="similar to EST AW341490 (NID:g6838116) hd10d04.x1"
     misc_feature    18040. .18510
                     /note="similar to EST AI807116 (NID:g5393682) we85g11.x1"
     misc_feature    18040. .18330
                     /note="similar to EST AI284657 (NID:g3022890) qu10g01.x1"
     misc_feature    18040. .18510
                     /note="similar to EST AI671887 (NID:g4851618) wb41c03.x1"
     misc_feature    18041. .18344
                     /note="similar to EST BB448631 (NID:g9309666)"
     misc_feature    18140. .18510
                     /note="similar to EST BE689032 (NID:g10076656) uw58g12.y1"
     misc_feature    18152. .18585

  Query Match            17.8%;  Score 636.6;  DB 9;  Length 159041;
  Best Local Similarity  98.6%;  Pred. No. 2.5e-128;
  Matches  642;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy     708 ggggaaatacgaatgtgctttaaagcgagcaagtattaaaattgtgactcctgactgggt 767
           || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   50682 GGAGAAATACGAATGTGCTTTAAAGCGAGCAAGTATTAAAATTGTGAC CTGACTGGGT 50623

Qy     768 tctggattgcgtatcagagaaaaccaaaaaggacgaagcattttatcatcctcgtctgat 827
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   50622 TCTGGATTGCGTATCAGAGAAAACCAAAAAGGACGAAGCATTTTATCATCCTCGTCTGAT 50563

Qy     828 tatttatgaagaggaagaagaggaagaggaagaggaggaggaagtagaaaatgaggaaca 887
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   50562 TATTTATGAAGAGGAAGAAGAGGAAGAGGAAGAGGAGGAGGAAGTAGAAAATGAGGAACA 50503

Qy     888 agattctcagaatgagggtagtacagatgagaagtcaagccctgccagctctcaagaagg 947
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   50502 AGATTCTC GAATGAGGGTAGTACAGATGAGAAGTCAAGCCCTGCCAGCTCTCAAGAAGG 50443

Qy     948 gtctccttcaggtgaccagcagttttcacctaaatccaacactgaaaaatctaaaqgqqa 1007
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   50442 GTCTCCTTCAGGTGACCAGCAGTTTTCACCTAAATCCAACACTGAAAAATCTAAAGGGGA 50383

Qy    1008 attaatgtttgatgattcttcagattcatcaccggaaaaacaggagagaaatttaaactg 1067
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   50382 ATTAATGTTTGATGATTCTTCAGATTCATCACCGGAAAAACAGGAGAGAAATTTAAACTG 50323

Qy    1068 gaccccggccgaagtcccacagttagctgcagcaaaacgcaggctgcctcagggaaagga 1127
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   50322 GACCCCGGCCGAAGTCCCACAGTTAGCTGCAGCAAAACGCAGGCTGCCTCAGGGAAAGGA 50263

Qy    1128 gcctgggttgattaacttgtgtgccaatgtcccacccgtcccaggtaacattttgccccc 1187
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   50262 GCCTGGGTTGATTAACTTGTGTGCCAATGTCCCACCCGTCCCAGGTAACATTTTGCCCCC 50203

Qy    1188 tgaggtccggggtaatttaatggctgctggacaaaacctccaaagttctgaaagatcaga 1247
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   50202 TGAGGTCCGGGGTAATTTAATGGCTGCTGGACAAAACCTCCAAAGTTCTGAAAGATCAGA 50143

Qy    1248 aatgatagctacctggagtccagctgtacggacactgaggaatattactaataatgctga 1307
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   50142 AATGATAGCTACCTGGAGTCCAGCTGTACGGACACTGAGGAATATTACTAATAATGCTGA 50084

Qy    1308 cattcagcagatgaaccggccatcaaatgtagcacatatcttacagactct 1358
           ||||||||||||||||||||||||||||||||||| |    ||  || | |
Db   50082 CATTCAGCAGATGAACCGGCCATCAAATGTAGCACATGTAAGTCACTCTTT 50032


RESULT   7
AC073077/c
LOCUS       AC073077               149592 bp    DNA     linear   HTG 17-NOV-2000
DEFINITION  Homo sapiens chromosome 7 clone RP11-269M19, WORKING DRAFT
            SEQUENCE, 10 unordered pieces.
ACCESSION   AC073077
```

```
VERSION     AC073077.5  GI:11192192
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 149592)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 149592)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JUN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Nov 17, 2000 this sequence version replaced gi:11054999.

            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            ---------- Project Information ----------
            Center project name: H_NH0269M19
            ---------- Summary Statistics ----------
            Sequencing vector: M13; 18%
            Sequencing vector: plasmid; 64%
            Chemistry: Dye-primer ET; 18% of reads
            Chemistry: Dye-terminator Big Dye; 64% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 142511 bases at least Q40
            Consensus quality: 144983 bases at least Q30
            Consensus quality: 146462 bases at least Q20
            Insert size: 167000; agarose-fp
            Insert size: 148692; sum-of-contigs
            Quality coverage: 5.83 in Q20 bases; agarose-fp
            Quality coverage: 5.97 in Q20 bases; sum-of-contigs
            ----------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     4209: contig of 4209 bp in length
            *    4210     4309: gap of unknown length
            *    4310    25449: contig of 21140 bp in length
            *   25450    25549: gap of unknown length
            *   25550    42417: contig of 16868 bp in length
            *   42418    42517: gap of unknown length
            *   42518    60534: contig of 18017 bp in length
            *   60535    60634: gap of unknown length
            *   60635    97012: contig of 36378 bp in length
            *   97013    97112: gap of unknown length
            *   97113   139971: contig of 42859 bp in length
            *  139972   140071: gap of unknown length
            *  140072   141186: contig of 1115 bp in length
            *  141187   141286: gap of unknown length
            *  141287   143081: contig of 1795 bp in length
            *  143082   143181: gap of unknown length
            *  143182   145560: contig of 2379 bp in length
            *  145561   145660: gap of unknown length
            *  145661   149592: contig of 3932 bp in length.
FEATURES             Location/Qualifiers
     source          1. .149592
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /clone="RP11-269M19"
     misc_feature    1. .4209
                     /note="assembly_name:Contig10
                     clone_end:T7
                     vector_side:left"
     misc_feature    4310. .25449
                     /note="assembly_name:Contig11"
     misc_feature    25550. .42417
                     /note="assembly_name:Contig12"
     misc_feature    42518. .60534
                     /note="assembly_name:Contig13
                     clone_end:SP6
                     vector_side:left"
     misc_feature    60635. .97012
                     /note="assembly_name:Contig14"
     misc_feature    97113. .139971
                     /note="assembly_name:Contig15"
     misc_feature    140072. .141186
                     /note="assembly_name:Contig6"
     misc_feature    141287. .143081
                     /note="assembly_name:Contig7"
     misc_feature    143182. .145560
                     /note="assembly_name:Contig8"
     misc_feature    145661. .149592
                     /note="assembly_name:Contig9"
BASE COUNT    40421 a  34230 c  33544 g  40490 t    907 others
ORIGIN


  Query Match            13.4%;  Score 479.2;  DB 2;  Length 149592;
  Best Local Similarity  98.4%;  Pred. No. 6.9e-94;
  Matches  484;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy   3079 ctggactatgaatcatataagtttaactgatggcgtctaggctgccgtgcatgtcgactc 3138
          ||  || |     ||||||||||||||||||||||||||||||||||||||||||||||
Db 111411 CTCAACAAGCCTACATATAAGTTTAACTGATGGCGTCTAGGCTGCCGTGCATGTCGACTC 111352

Qy   3139 ctgcggtccggggctggctgtctggctggcgaggagctgctgcgcttccttcacatgctc 3198
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 111351 CTGCGGTGCGGGGCTGGCTGTCTGGCTGGCGAGGAGCTGCTGCGCTTCCTTCACATGCTC 111292

Qy   3199 ttgttttccagctgctttcctgggggatcagactgtgaagcaggaagacagatataataa 3258
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 111291 TTGTTTTCCAGCTGCTTTCCTGGGGGATCAGACTGTGAAGCAGGAAGACAGATATAATAA 111232

Qy   3259 atatactgcatctttttaagatgtgcaattttattctgaggaaacataaattatgttttg 3318
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 111231 ATATACTGCATCTTTTTAAGATGTGCAATTTTATTCTGAGGAAACATAAATTATGTTTTG 111172

Qy   3319 tattatatgactttaagagcccacattaggttttatgattcatttgccaggtttttaaat 3378
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 111171 TATTATATGACTTTAAGAGCCCACATTAGGTTTTATGATTCATTTGCCAGGTTTTTAAAT 111112

Qy   3379 gttttcacaaaactgttacgggacttcaactagaaataaaatggtgtaaataaagacctt 3438
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 111111 GTTTTCACAAAACTGTTACGGGACTTCAACTAGAAATAAAATGGTGTAAATAAAGACCTT 111052

Qy   3439 gctatctctaaattatggatgttaaagatttgaaatgttttgtactttgattatttttat 3498
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 111051 GCTATCTCTAAATTATGGATGTTAAAGATTTGAAATGTTTTGTACTTTGATTATTTTTAT 110992

Qy   3499 ttcttatactctgttttcttttatattgatatcttgcccacattttaaataaatgtactt 3558
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110991 TTCT ATACTCTGTTTTCTTTTATATTGATATCTTGCCCACATTTTAAATAAATGTACTT 110932

Qy   3559 ttgaacttagaa 3570
          ||||||||||||
Db 110931 TTGAACTTAGAA 110920


RESULT   8
G30194/c
LOCUS       G30194                   438 bp    DNA     linear   STS 05-OCT-1996
DEFINITION  human STS SHGC-36464, sequence tagged site.
ACCESSION   G30194
VERSION     G30194.1  GI:1593745
```

```
KEYWORDS    STS; STS sequence; primer; sequence tagged site.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 438)
  AUTHORS   Myers,R.M.
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Richard M. Myers
            Stanford Human Genome Center (SHGC)
            Stanford University School of Medicine
            Department of Genetics, M-344, Stanford, CA 94305, USA
            Tel: 4157259687
            Fax: 4157259689
            Email: myers@shgc.stanford.edu

            Primer A: TATTTAAAATGTGGGCAAGATATCA
            Primer B: TGGTGTAAATAAAGACCTTGCTATC
            STS size: 130
            PCR Profile:
                    Initial incubation: 94 degrees C for 90 seconds

                    Denaturation:       94 degrees C for 15 seconds
                    Annealing:          62 degrees C for 23 seconds
                    Polymerization:     72 degrees C for 30 seconds
                    PCR Cycles:         30
                    Thermal Cycler:     Perkin Elmer 9600
            Protocol:
                    Template:       25 ng
                    Primer:         each 1 uM
                    dNTPs:          each 200 uM
                    Taq Polymerase: 0.05 units/ul
                    Total Vol:      10 ul

            Buffer:
                    MgCl2:          2.5 mM
                    KCl:            50  mM
                    Tris-HCl:       20  mM
                    pH:             8.3

            Prepared with primer pairs provided by Sandoz, derived from T88893
            -- Washington University/Merck EST sequence.
FEATURES            Location/Qualifiers
     source         1. .438
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /map="7"
     STS            20. .149
     primer_bind    20. .44
     primer_bind    complement(125. .149)
BASE COUNT      155 a     82 c     77 g    117 t      7 others
ORIGIN
```

```
  Query Match           10.1%;  Score 361.4;  DB 11;  Length 438;
  Best Local Similarity  94.4%;  Pred. No. 1.7e-68;
  Matches  405;  Conservative    0;  Mismatches   20;  Indels    4;  Gaps    3;

Qy   3138 cctgcggtgcggggctggctgtctggctggcgaggagctgctgcgctt-ccttcacatgc 3196
          || |||||||||||| ||||| | ||| |||||||     | |||||| ||||||||||
Db    435 CCCGCGGTGCGGGGCNGGCTGCCGGGCNGCGCGAGGGAGCCCNGCGCTTCCCTTCACATGG 376

Qy   3197 tcttgtttt-ccagctgctttcctgggggat--cagactgtgaagcaggaagacagatat 3253
          | |||||| |||||||||| ||  ||||    ||||||||||||||||||||||||||
Db    375 CCCTGTTTTCCCAGCTGCTTNCCCTGGGGGATCCAGACTGTGAAGCAGGAAGACAGATAT 316

Qy   3254 aataaatatactgcatctttttaagatgtgcaattttattctgaggaaacataaattatg 3313
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    315 AATAAATATACTGCATCTTTTTAAGATGTGCAATTTTATTCTGAGGAAACATAAATTATG 256

Qy   3314 ttttgtattatatgactttaagagcccacattaggttttatgattcatttgccaggtttt 3373
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    255 TTTTGTATTATATGACTTTAAGAGCCCACATTAGGTTTTATGATTCATTTGCCAGGTTTT 196

Qy   3374 taaatgttttcacaaaactgttacgggacttcaactagaaataaaatggtgtaaataaag 3433
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    195 TAAATGTTTTCACAAAACTGTTACGGGACTTCAACTAGAAATAAAATGGTGTAAATAAAG 136

Qy   3434 accttgctatctctaaattatggatgttaaagatttgaaatgttttgtactttgattatt 3493
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    135 ACCTTGCTATCTCTAAATTATGGATGTTAAAGATTTGAAATGTTTTGTACTTTGATTATT 76

Qy   3494 tttatttcttatactctgttttcttttatattgatatcttgcccacattttaaataaatg 3553
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     75 TTTATTTCTTATACTCTGTTTTCTTTTATATTGATATCTTGCCCACATTTTAAATAAATG 16

Qy   3554 tacttttga 3562
          ||||||| |
Db     15 TACTTTTCA 7
```

```
RESULT   9
AC094546/c
LOCUS       AC094546              120887 bp    DNA     linear   HTG 20-DEC-2001
DEFINITION  Rattus norvegicus clone CH230-4I21, *** SEQUENCING IN PROGRESS ***,
            74 unordered pieces.
ACCESSION   AC094546
VERSION     AC094546.2  GI:17941297
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 120887)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
```

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 120887)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (15-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Dec 20, 2001 this sequence version replaced gi:15624381.
---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GAWE
Center clone name: CH230-4I21
---------- Summary Statistics
Assembly program: Phrap; version 0.990329First call to findPhrapList
Consensus quality: 81355 bases at least Q40
Consensus quality: 89305 bases at least Q30
Consensus quality: 95167 bases at least Q20
Estimated insert size: 51481; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 0.6x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 74 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

*	1	2144: contig of 2144 bp in length
*	2145	2244: gap of unknown length
*	2245	5011: contig of 2767 bp in length
*	5012	5111: gap of unknown length
*	5112	7135: contig of 2024 bp in length
*	7136	7235: gap of unknown length
*	7236	9223: contig of 1988 bp in length
*	9224	9323: gap of unknown length
*	9324	11509: contig of 2186 bp in length
*	11510	11609: gap of unknown length
*	11610	13474: contig of 1865 bp in length
*	13475	13574: gap of unknown length
*	13575	15662: contig of 2088 bp in length
*	15663	15762: gap of unknown length
*	15763	17929: contig of 2167 bp in length
*	17930	18029: gap of unknown length
*	18030	20836: contig of 2807 bp in length
*	20837	20936: gap of unknown length
*	20937	22986: contig of 2050 bp in length
*	22987	23086: gap of unknown length
*	23087	25197: contig of 2111 bp in length
*	25198	25297: gap of unknown length
*	25298	26932: contig of 1635 bp in length
*	26933	27032: gap of unknown length
*	27033	29184: contig of 2152 bp in length
*	29185	29284: gap of unknown length
*	29285	31122: contig of 1838 bp in length
*	31123	31222: gap of unknown length
*	31223	33089: contig of 1867 bp in length
*	33090	33189: gap of unknown length
*	33190	34621: contig of 1432 bp in length
*	34622	34721: gap of unknown length
*	34722	36131: contig of 1410 bp in length
*	36132	36231: gap of unknown length
*	36232	37852: contig of 1621 bp in length
*	37853	37952: gap of unknown length
*	37953	39448: contig of 1496 bp in length
*	39449	39548: gap of unknown length
*	39549	41342: contig of 1794 bp in length
*	41343	41442: gap of unknown length
*	41443	43126: contig of 1684 bp in length
*	43127	43226: gap of unknown length
*	43227	44363: contig of 1137 bp in length
*	44364	44463: gap of unknown length
*	44464	45992: contig of 1529 bp in length
*	45993	46092: gap of unknown length
*	46093	47391: contig of 1299 bp in length
*	47392	47491: gap of unknown length
*	47492	49188: contig of 1697 bp in length
*	49189	49288: gap of unknown length
*	49289	50333: contig of 1045 bp in length
*	50334	50433: gap of unknown length
*	50434	52192: contig of 1759 bp in length
*	52193	52292: gap of unknown length
*	52293	54222: contig of 1930 bp in length
*	54223	54322: gap of unknown length
*	54323	55865: contig of 1543 bp in length
*	55866	55965: gap of unknown length
*	55966	56998: contig of 1033 bp in length
*	56999	57098: gap of unknown length
*	57099	58889: contig of 1791 bp in length
*	58890	58989: gap of unknown length
*	58990	61298: contig of 2309 bp in length
*	61299	61398: gap of unknown length
*	61399	63560: contig of 2162 bp in length
*	63561	63660: gap of unknown length
*	63661	65400: contig of 1740 bp in length
*	65401	65500: gap of unknown length
*	65501	66539: contig of 1039 bp in length
*	66540	66639: gap of unknown length
*	66640	67683: contig of 1044 bp in length
*	67684	67783: gap of unknown length
*	67784	68898: contig of 1115 bp in length
*	68899	68998: gap of unknown length
*	68999	70929: contig of 1931 bp in length
*	70930	71029: gap of unknown length
*	71030	72545: contig of 1516 bp in length
*	72546	72645: gap of unknown length
*	72646	73833: contig of 1188 bp in length
*	73834	73933: gap of unknown length
*	73934	75016: contig of 1083 bp in length
*	75017	75116: gap of unknown length
*	75117	76334: contig of 1218 bp in length
*	76335	76434: gap of unknown length
*	76435	78199: contig of 1765 bp in length
*	78200	78299: gap of unknown length
*	78300	80037: contig of 1738 bp in length
*	80038	80137: gap of unknown length
*	80138	81196: contig of 1059 bp in length
*	81197	81296: gap of unknown length
*	81297	82762: contig of 1466 bp in length
*	82763	82862: gap of unknown length
*	82863	84079: contig of 1217 bp in length
*	84080	84179: gap of unknown length
*	84180	85257: contig of 1078 bp in length
*	85258	85357: gap of unknown length
*	85358	86571: contig of 1214 bp in length
*	86572	86671: gap of unknown length
*	86672	87945: contig of 1274 bp in length
*	87946	88045: gap of unknown length
*	88046	89172: contig of 1127 bp in length
*	89173	89272: gap of unknown length
*	89273	90314: contig of 1042 bp in length
*	90315	90414: gap of unknown length
*	90415	91836: contig of 1422 bp in length
*	91837	91936: gap of unknown length
*	91937	93103: contig of 1167 bp in length
*	93104	93203: gap of unknown length
*	93204	94356: contig of 1153 bp in length
*	94357	94456: gap of unknown length

```
          *    94457    95913: contig of 1457 bp in length
          *    95914    96013: gap of unknown length

  Query Match             9.9%;  Score 353.4;  DB 2;  Length 120887;
  Best Local Similarity   74.0%;  Pred. No. 2.2e-66;
  Matches  492;  Conservative    0;  Mismatches  161;  Indels   12;  Gaps    3;

Qy    682 catttgattgttccagagccaaaggggggggaaatacgaatgtgctttaaagcgagcaagt 741
          |||| |  |||         |        || ||| |||||  ||||| | ||| || ||||
Db   8977 CATTAGTATGTTATGCTTCACTCCTAGGAGAAGTACGAGCGTGCTGTCAAGAGAACAAGC 8918

Qy    742 attaaaattgtgactcctgactgggttctggattgcgtatcagagaaaaccaaaaaggac 801
          ||||| ||||| || |||||||||||| ||||| || || || ||||| |  |  |||||
Db   8917 ATTAAGATTGTTACCCCTGACTGGGTCCTGGACTGTGTGTCTGAGAAGAGGAGGAAGGAT 8858

Qy    802 gaagcattttatcatcctcgtctgattatttatgaagaggaagaagaggaagaggaagag 861
          ||||| | ||| ||||| |||||||| || ||||| ||||||      || ||||| |||
Db   8857 GAAGCTTCTTACCATCCACGTCTGATCATCTATGAGGAGGAA------GAGGAGGAGGAG 8804

Qy    862 gaggaggaagtagaaaatgaggaacaagattctcaqaatgagggtagtacagatgagaag 921
          ||||| || | ||| || ||||| || || |||||||| |||||||| ||   ||| |||
Db   8803 GAGGAAGAGGGAGACAACGAGGAGCAGGACTCTCAGAACGAGGGTAGCAC---TGACAAG 8747

Qy    922 tcaagccctgccagctctcaagaagggtctccttcaggtgaccagcagttttcacctaaa 981
          || ||  | |||||||||   || |  || |||   | |||||||| ||  || || ||
Db   8746 TCCAGTGCAGCCAGCTCTGCGGAGGCCTCCCCT---GCTGACCAGCCGTGCTCTCCCAAG 8690

Qy    982 tccaacactgaaaaatctaaaggggaattaatgtttgatgattcttcagattcatcaccg 1041
           |||   ||||   ||| |||||||| || ||||||||||| ||||| || ||||| ||
Db   8689 GCCAGGGCTGAGGTATCCAAAGGGGAGTTGATGTTTGATGACTCTTCTGACTCATCCCCT 8630

Qy   1042 gaaaaacaggagagaaatttaaactggaccccggccgaagtcccacagttagctgcagca 1101
          ||||| ||||| | |  || |||||| |||| || |||| ||| |   |     ||||
Db   8629 GAAAAGCAGGAGCGCAGCTTGAACTGCGCCCCCGCTGAAGCCCCGCCACTCAGCACAGCC 8570

Qy   1102 aaacgcaggctgcctcagggaaaggagcctgggttgattaacttgtgtgccaatgtccca 1161
          | | ||||||||||||||||||| |  || ||  | |||||| ||||||||||||||||
Db   8569 CAGCACAGGCTGCCTCAGGGAAAAGGACCGGGACTCATTAACCTGTGTGCCAATGTCCCA 8510

Qy   1162 cccgtcccaggtaacattttgccccctgaggtccggggtaatttaatggctgctggacaa 1221
          || ||||||||  |||||| |||| |||||  | |||||||| || ||||| |||||||||
Db   8509 CCTGTCCCAGGGGACATTCTGCCTCCTGACATGCGGGGTAACTTGATGGCCGCTGGACAA 8450

Qy   1222 aacctccaaagttctgaaagatcaqaaatgataqctacctggagtccagctgtacggaca 1281
          |||||||| ||||||||  |  | |||||| |||  |||||||| |||||||| |||||
Db   8449 AACCTCCAGAGTTCTGAGCGGCCGGAAATGTTAGGGACCTGGAGCCCAGCTGTCCGGACG 8390

Qy   1282 ctgaggaatattactaataatgctgacattcagcagatgaaccggccatcaaatgtagca 1341
          || | || ||||| || || ||||| || ||||| || | |||||||| ||||| |||
Db   8389 TTGCGAAACATTACAAACAACGCTGATATCCAGCAAATCAGCCGGCCATCCAATGTTGCA 8330

Qy   1342 catat 1346
          ||| |
Db   8329 CATGT 8325


RESULT  10
AC074040
LOCUS       AC074040               178419 bp    DNA     linear   HTG 13-JUL-2000
DEFINITION  Mus musculus chromosome 5 clone RP23-130M12 strain C57BL6/J,
            WORKING DRAFT SEQUENCE, 24 unordered pieces.
ACCESSION   AC074040
VERSION     AC074040.1  GI:9087212
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 178419)
  AUTHORS   Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
            Bouffard,G.G., Dietrich,N.L., Eagle,W.O., Gupta,J., Ho,S.-L.,
            Huang,M.C., Idol,J., Lee-Lin,S.-Q., Maduro,Q.L., Maduro,V.B.,
            Mastrian,S.D., McCloskey,J.C., Ojodu,M.A., Pearson,R.,
            Stantripop,S., Summers,T.J., Thomas,J.W., Thomas,P.J.,
            Tiongson,E.E., Touchman,J.W., Tran,J.T., Vogt,J.L., Walker,M.A.,
            Wetherby,K.D. and Green,E.D.
  TITLE     NISC Mouse Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 178419)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-2000) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     -------------- Genome Center
            Center: NIH Intramural Sequencing Center
            Center code: NISC
            Web site: http://www.nisc.nih.gov
            Contact: nisc_mouse@nhgri.nih.gov
            -------------- Project Information
            Center project name: tk
            Center clone name: 130M12
            -------------- Summary Statistics
            Sequencing vector: plasmid; n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 164732 bases at least Q40
            Consensus quality: 168102 bases at least Q30
            Consensus quality: 170232 bases at least Q20
            Insert size: 183000; agarose-fp
            Insert size: 177000; pulse-field-gel
            Insert size: 176119; sum-of-contigs
            Quality coverage: 4.22x in Q20 bases; agarose-fp
            Quality coverage: 4.37x in Q20 bases; pulse-field-gel
            Quality coverage: 4.39x in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 24 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2335: contig of 2335 bp in length
            *     2336     2435: gap of unknown length
            *     2436     4708: contig of 2273 bp in length
            *     4709     4808: gap of unknown length
            *     4809     7741: contig of 2933 bp in length
            *     7742     7841: gap of unknown length
            *     7842    11301: contig of 3460 bp in length
            *    11302    11401: gap of unknown length
            *    11402    16198: contig of 4797 bp in length
            *    16199    16298: gap of unknown length
            *    16299    21111: contig of 4813 bp in length
            *    21112    21211: gap of unknown length
            *    21212    26078: contig of 4867 bp in length
            *    26079    26178: gap of unknown length
            *    26179    31586: contig of 5408 bp in length
            *    31587    31686: gap of unknown length
            *    31687    38653: contig of 6967 bp in length
            *    38654    38753: gap of unknown length
            *    38754    44405: contig of 5652 bp in length
            *    44406    44505: gap of unknown length
            *    44506    50909: contig of 6404 bp in length
            *    50910    51009: gap of unknown length
            *    51010    56618: contig of 5609 bp in length
            *    56619    56718: gap of unknown length
            *    56719    65690: contig of 8972 bp in length
            *    65691    65790: gap of unknown length
            *    65791    75015: contig of 9225 bp in length
            *    75016    75115: gap of unknown length
            *    75116    82126: contig of 7011 bp in length
            *    82127    82226: gap of unknown length
            *    82227    89936: contig of 7710 bp in length
            *    89937    90036: gap of unknown length
            *    90037    99253: contig of 9217 bp in length
```

```
            *    99254    99353: gap of unknown length
            *    99354   110325: contig of 10972 bp in length
            *   110326   110425: gap of unknown length
            *   110426   119465: contig of 9040 bp in length
            *   119466   119565: gap of unknown length
            *   119566   128843: contig of 9278 bp in length
            *   128844   128943: gap of unknown length
            *   128944   138762: contig of 9819 bp in length
            *   138763   138862: gap of unknown length
            *   138863   152091: contig of 13229 bp in length
            *   152092   152191: gap of unknown length
            *   152192   164766: contig of 12575 bp in length
            *   164767   164866: gap of unknown length
            *   164867   178419: contig of 13553 bp in length.
FEATURES             Location/Qualifiers
     source          1. .178419
                     /organism="Mus musculus"
                     /strain="C57BL6/J"
                     /db_xref="taxon:10090"
                     /chromosome="5"
                     /clone="RP23-130M12"
                     /clone_lib="RPCI mouse BAC library 23"
     misc_feature    1. .2335
                     /note="assembly_fragment"
     misc_feature    2436. .4708
                     /note="assembly_fragment"
     misc_feature    4809. .7741
                     /note="assembly_fragment"
     misc_feature    7842. .11301
                     /note="assembly_fragment"
     misc_feature    11402. .16198
                     /note="assembly_fragment"
     misc_feature    16299. .21111
                     /note="assembly_fragment"
     misc_feature    21212. .26078
                     /note="assembly_fragment"
     misc_feature    26179. .31586
                     /note="assembly_fragment"
     misc_feature    31687. .38653
                     /note="assembly_fragment"
     misc_feature    38754. .44405
                     /note="assembly_fragment"
     misc_feature    44506. .50909
                     /note="assembly_fragment"
     misc_feature    51010. .56618
                     /note="assembly_fragment"
     misc_feature    56719. .65690
                     /note="assembly_fragment"
     misc_feature    65791. .75015
                     /note="assembly_fragment"
     misc_feature    75116. .82126
                     /note="assembly_fragment"
     misc_feature    82227. .89936
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    90037. .99253
                     /note="assembly_fragment"
     misc_feature    99354. .110325
                     /note="assembly_fragment"
     misc_feature    110426. .119465
                     /note="assembly_fragment"
     misc_feature    119566. .128843
                     /note="assembly_fragment"
     misc_feature    128944. .138762
                     /note="assembly_fragment"
     misc_feature    138863. .152091
                     /note="assembly_fragment"
     misc_feature    152192. .164766
                     /note="assembly_fragment"
     misc_feature    164867. .178419
                     /note="assembly_fragment"
BASE COUNT    48439 a   40696 c   40698 g   46277 t    2309 others
```

```
ORIGIN


  Query Match             9.7%;  Score 348.6;  DB 2;  Length 178419;
  Best Local Similarity   75.1%;  Pred. No. 2.7e-65;
  Matches  480;  Conservative    0;  Mismatches  144;  Indels   15;  Gaps    3;

Qy     708 ggggaaatacgaatgtgctttaaagcgagcaagtattaaaattgtgactcctgactgggt 767
           || ||||||||| |||| | ||  || |||||||||| ||||| || || ||||||||
Db   97839 GGAGAAATACGAGCGTGCTGTCAAAAGAACAAGTATTAAGATTGTCACCCCGGACTGGGT 97898

Qy     768 tctggattgcgtatcagagaaaaccaaaaaggacgaagcattttatcatcctcgtctgat 827
           ||||| || || || ||||| |  | ||||| ||||| ||||| ||||| || ||||||
Db   97899 CCTGGACTGTGTGTCTGAGAAGAGGAGGAAGGATGAAGCGTTTTACCATCCACGCCTGAT 97958

Qy     828 tatttatgaagaggaagaagaggaagaggaagaggaggaggaagtagaaaatgaggaaca 887
           || ||||| ||||| |||||||| |||||||||          | ||| || ||||| ||
Db   97959 CATCTATGAGGAGGAGGAAGAGGAGGAGGAAGAG---------GGAGACAACGAGGAGCA 98009

Qy     888 agattctcagaatgagggtagtacagatgagaagtcaagccctgccagctctcaagaagg 947
           || || ||||| ||||| || ||   |||||||||| ||    |||||||||  || |
Db   98010 GGACTCCCAGAACGAGGGCAGCAC---TGAGAAGTCCAGTGTCGCCAGCTCTGCAGTGGC 98066

Qy     948 gtctccttcaggtgaccagcagttttcacctaaatccaacactgaaaaatctaaaggaga 1007
           || |||   | ||| |||| ||  || || ||  |||   ||||    || ||||| ||
Db   98067 CTCCCCT---GCTGAGCAGCCGTGCTCTCCCAAGCCCAGGGCTGAG.TGTCCAAAGGAGA 98123

Qy    1008 attaatgtttgatgattcttcagattcatcaccggaaaaacaggagagaaatttaaactg 1067
           || |||||||||| |||| || |||| || |||| |||||| | |  || |||||
Db   98124 GTTGATGTTTGATGACTCTTCTGACTCATCCCCTGAAAAGCAGGAGCGCAGCTTGAACTG 98183

Qy    1068 gaccccggccgaagtcccacagttagctgcagcaaaacgcaggctgcctcagggaaagga 1127
           | |||| || |||| |||||   |     ||||  | ||||||||||||||||||| |
Db   98184 GGCCCCTGCTGAAGCCCCACCACTCAACACAGCCCAGCGCAGGCTGCCTCAGGGAAAAGG 98243

Qy    1128 gcctgggttgattaacttgtgtgccaatgtcccacccgtcccaggtaacattttgccccc 1187
           ||| ||  | |||||| ||||||||||||||||||| || |||||  |||| |||| ||
Db   98244 GCCAGGA.TCATTAACCTGTGTGCCAATGTCCCACCTGTTCCAGGGGACATTCTGCCTCC 98303

Qy    1188 tgaggtccggggtaatttaatggctgctggacaaaacctccaaagttctgaaagatcaga 1247
           |||  | |||||||| || |||||  |||||||||||||||| | ||||||  | |||||
Db   98304 TGACATGCGGGGTAACTTGATGGCCCCTGGACAAAACCTCCAGAATTCTGAGCGGTCAGA 98363

Qy    1248 aatgatagctacctggagtccagctgtacggacactgaggaatattactaataatgctga 1307
           |||  |||  |||||||| |||||||| |||||  || | || ||||| || ||||||||
Db   98364 AATACTAGGGACCTGGAGCCCAGCTGTCCGGACGTTGCGAAACATTACAAACAATGCTGA 98423

Qy    1308 cattcagcagatgaaccggccatcaaatgtagcacatat 1346
           || |||||||| |||||||||| ||||| |||||| |
Db   98424 TATCCAGCAGATCAACCGGCCATCCAATGTTGCACATGT 98462


RESULT  11
AC079217
LOCUS       AC079217               177205 bp    DNA     linear   HTG 24-AUG-2000
DEFINITION  Mus musculus chromosome 5 clone RP23-381D23 strain C57BL6/J,
            WORKING DRAFT SEQUENCE, 20 unordered pieces.
ACCESSION   AC079217
VERSION     AC079217.1  GI:9887697
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 177205)
  AUTHORS   Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
            Bouffard,G.G., Dietrich,N.L., Eagle,W.O., Gupta,J., Ho,S.-L.,
            Idol,J., Lee-Lin,S.-Q., Legaspi,R., Lim,M., Maduro,Q.L.,
            Maduro,V.B., Mastrian,S.D., McCloskey,J.C., McDowell,J.,
            Pearson,R., Stantripop,S., Summers,T.J., Thomas,J.W., Thomas,P.J.,
            Tiongson,E.E., Touchman,J.W., Tran,J.T., Vogt,J.L., Walker,M.A.,
            Wetherby,K.D. and Green,E.D.
  TITLE     NISC Mouse Sequencing Initiative
```

```
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 177205)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-AUG-2000) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     ---------- Genome Center
            Center: NIH Intramural Sequencing Center
            Center code: NISC
            Web site: http://www.nisc.nih.gov
            Contact: nisc_mouse@nhgri.nih.gov
            ---------- Project Information
            Center project name: tl
            Center clone name: 381D23
            ---------- Summary Statistics
            Sequencing vector: plasmid; n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 164801 bases at least Q40
            Consensus quality: 167308 bases at least Q30
            Consensus quality: 168762 bases at least Q20
            Insert size: 209000; agarose-fp
            Insert size: 188000; pulse-field-gel
            Insert size: 175305; sum-of-contigs
            Quality coverage: 4.10x in Q20 bases; agarose-fp
            Quality coverage: 4.56x in Q20 bases; pulse-field-gel
            Quality coverage: 4.89x in Q20 bases; sum-of-contigs
            ----------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 20 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     3158: contig of 3158 bp in length
            *    3159     3258: gap of unknown length
            *    3259     6522: contig of 3264 bp in length
            *    6523     6622: gap of unknown length
            *    6623    10277: contig of 3655 bp in length
            *   10278    10377: gap of unknown length
            *   10378    14299: contig of 3922 bp in length
            *   14300    14399: gap of unknown length
            *   14400    18285: contig of 3886 bp in length
            *   18286    18385: gap of unknown length
            *   18386    24955: contig of 6570 bp in length
            *   24956    25055: gap of unknown length
            *   25056    30203: contig of 5148 bp in length
            *   30204    30303: gap of unknown length
            *   30304    36272: contig of 5969 bp in length
            *   36273    36372: gap of unknown length
            *   36373    40990: contig of 4618 bp in length
            *   40991    41090: gap of unknown length
            *   41091    49862: contig of 8772 bp in length
            *   49863    49962: gap of unknown length
            *   49963    55586: contig of 5624 bp in length
            *   55587    55686: gap of unknown length
            *   55687    65769: contig of 10083 bp in length
            *   65770    65869: gap of unknown length
            *   65870    76152: contig of 10283 bp in length
            *   76153    76252: gap of unknown length
            *   76253    87685: contig of 11433 bp in length
            *   87686    87785: gap of unknown length
            *   87786    97230: contig of 9445 bp in length
            *   97231    97330: gap of unknown length
            *   97331   112484: contig of 15154 bp in length
            *  112485   112584: gap of unknown length
            *  112585   127931: contig of 15347 bp in length
            *  127932   128031: gap of unknown length
            *  128032   143545: contig of 15514 bp in length
            *  143546   143645: gap of unknown length
            *  143646   159112: contig of 15467 bp in length
            *  159113   159212: gap of unknown length
            *  159213   177205: contig of 17993 bp in length.
FEATURES             Location/Qualifiers
     source          1. .177205
                     /organism="Mus musculus"
                     /strain="C57BL6/J"
                     /db_xref="taxon:10090"
                     /chromosome="5"
                     /clone="RP23-381D23"
                     /clone_lib="RPCI mouse BAC library 23"
     misc_feature    1. .3158
                     /note="assembly_fragment"
     misc_feature    3259. .6522
                     /note="assembly_fragment"
     misc_feature    6623. .10277
                     /note="assembly_fragment"
     misc_feature    10378. .14299
                     /note="assembly_fragment"
     misc_feature    14400. .18285
                     /note="assembly_fragment"
     misc_feature    18386. .24955
                     /note="assembly_fragment"
     misc_feature    25056. .30203
                     /note="assembly_fragment"
     misc_feature    30304. .36272
                     /note="assembly_fragment"
     misc_feature    36373. .40990
                     /note="assembly_fragment"
     misc_feature    41091. .49862
                     /note="assembly_fragment"
     misc_feature    49963. .55586
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    55687. .65769
                     /note="assembly_fragment"
     misc_feature    65870. .76152
                     /note="assembly_fragment"
     misc_feature    76253. .87685
                     /note="assembly_fragment"
     misc_feature    87786. .97230
                     /note="assembly_fragment"
     misc_feature    97331. .112484
                     /note="assembly_fragment"
     misc_feature    112585. .127931
                     /note="assembly_fragment"
     misc_feature    128032. .143545
                     /note="assembly_fragment"
     misc_feature    143646. .159112
                     /note="assembly_fragment"
     misc_feature    159213. .177205
                     /note="assembly_fragment"
BASE COUNT    49012 a   39764 c   38186 g   48267 t    1976 others
ORIGIN


  Query Match             9.3%;  Score 334.4;  DB 2;  Length 177205;
  Best Local Similarity   74.7%;  Pred. No. 3.4e-62;
  Matches  478;  Conservative    0;  Mismatches  146;  Indels   16;  Gaps    4;

Qy     708 ggggaaatacgaatgtgctttaaagcgagcaagtattaaaattgtgactcctgactgggt 767
           || |||||||||  ||||| | ||  || |||||||||| ||||| || || |||||||
Db   73083 GGAGAAATACGAGCGTGCTGTCAAAAGAACAAGTATTAAGATTGTCACCCCGGACTGGGG 73142

Qy     768 t-ctggattgcgtatcagagaaaaccaaaaaggacgaagcattttatcatcctcgtctga 826
           | ||| | || || || ||||| |  | |||| ||||| ||||| ||||| || ||||
Db   73143 CCCTGAACTGTGTGTCTGAGAAGAGGAGGAAGGATGAAGCGTTTTACCATCCACGCCTGA 73202

Qy     827 ttatttatgaagaggaagaagaggaagaggaagaggaggaggaagtagaaaatgaggaac 886
           | || ||||| ||||| |||||||| |||||||||          | ||| || |||||
Db   73203 TCATCTATGAGGAGGAGGAAGAGGAGGAGGAAGAG---------GGAGACAACGAGGAGC 73253

Qy     887 aagattctcagaatgagggtagtacagatgagaagtcaagccctgccagctctcaagaag 946
```

```
Db  73254 AGGACTCCCAGAACGAGGGCAGCAC---TGAGAAGTCCAGTGTCGCCAGCTCTGCAGTGG 73310

Qy    947 qgtctccttcaggtgaccagcagtttttcacctaaatccaacactgaaaaatctaaagggg 1006
Db  73311 CCTCCCCT---GCTGAGCAGCCGTGCTCTCCCAAGCCCAGGGCTGAGGTGTCCAAAGGAG 73367

Qy   1007 aattaatgtttgatgattcttcagattcatcaccggaaaaacaggagagaaatttaaact 1066
Db  73368 AGTTGATGTTTGATGACTCTTCTGACTCATCCCCTGAAAAGCAGGAGCGCAGCTTGAACT 73427

Qy   1067 ggacccggccgaagtcccacagttagctgcagcaaaacgcaggctgcctcagggaaagg 1126
Db  73428 GGGCCCCTGCTGAAGCCCCACCACTCAACACAGCCCAGCGCAGGCTGCCTCAGGGAAAAG 73487

Qy   1127 agcctgggttgattaacttgtgtgccaatgtcccacccgtcccaggtaacattttgcccc 1186
Db  73488 GGCCAGGACTCATTAACCTGTGTGCCAATGTCCCACCTGTTCCAGGGGACATTCTGCCTC 73547

Qy   1187 ctgaggtccgggggtaatttaatggctgctggacaaaacctccaaagttctgaaagatcag 1246
Db  73548 CTGACATGCGGGGTAACTTGATGGCCCCTGGACAAAACCTCCAGAATTCTGAGCGGTCAG 73607

Qy   1247 aaatgatagctacctggagtccagctgtacggacactgaggaatattactaataatgctg 1306
Db  73608 AAATACTAGGGACCTGGAGCCCAGCTGTCCGGACGTTGCGAAACATTACAAACAATGCTG 73667

Qy   1307 acattcagcagatgaaccggccatcaaatgtagcacatat 1346
Db  73668 ATATCCAGCAGATCAACCGGCCATCCAATGTTGCACATGT 73707
```

RESULT 12
HSU23862/c
LOCUS HSU23862 633 bp mRNA linear PRI 24-JUL-2001
DEFINITION Human clone mcag32 chromosome 7 CTG repeat region.
ACCESSION U23862
VERSION U23862.1 GI:777403
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 633)
AUTHORS Yamagata,K., Takeda,J., Menzel,S., Chen,X., Eng,S., Lim,L.R., Concannon,P., Hanis,C.L., Spielman,R.S., Cox,N.J. and Bell,G.I.
TITLE Searching for NIDDM susceptibility genes: studies of genes with triplet repeats expressed in skeletal muscle
JOURNAL Diabetologia 39 (6), 725-730 (1996)
MEDLINE 96375458
PUBMED 8781769
REFERENCE 2 (bases 1 to 633)
AUTHORS Bell,G.I.
TITLE Direct Submission
JOURNAL Submitted (31-MAR-1995) Graeme I. Bell, Howard Hughes Medical Institute, University of Chicago, 5841 South Maryland Avenue, Chicago, IL 60637, USA
FEATURES Location/Qualifiers
source 1. .633
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/clone="mcag32"
/sex="female"
/tissue_type="skeletal muscle"
/clone_lib="Stratagene catalogue no. 936215"
/dev_stage="adult"
repeat_region 356. .433
/rpt_family="CTG"
BASE COUNT 107 a 152 c 207 g 167 t
ORIGIN

```
Query Match 8.6%; Score 308.2; DB 9; Length 633;
Best Local Similarity 82.0%; Pred. No. 7.9e-57;
Matches 355; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

Qy   1465 cacatgctacagcagcagcagcaggcccagcagcagcagcagcagcacccggttttacac 1524
Db    433 CAGCTCCAGCAGCAGCAGCAACAGCAGATTCAGCAGCAGCAGCTCCAGCGCATGCACCAG 374

Qy   1525 cttcagccccagcagataatgcagctccagcagcagcagcagcagcagatctctcagcaa 1584
Db    373 CAGCAGCAGCAGCAGCAGATGCAAAGTCAGACAGCGCCACACTTGAGTCAGACGTCACAG 314

Qy   1585 ccttaccccagcagccgccgcatccatttttcacagcagcagcagcagcaacagcagccca 1644
Db    313 GCGCTGCAGCATCAGGTTCCACCTCAGCAGCCCCCGCAGCAGCAGCAACAGCAGCCA 254

Qy   1645 ccaccatcgcctcagcagcatcagctttttggacatgatccagcagtggagaattccagaa 1704
Db    253 CCACCATCGCCTCAGCAGCATCAGCTTTTTGGACATGATCCAGCAGTGGAGATTCCAGAA 194

Qy   1705 gaaggcttcttattgggatgtgtgtttgcaattgcggattatccagagcagatgtctgat 1764
Db    193 GAAGGCTTCTTATTGGGATGTGTGTTTGCAATTGCGGATTATCCAGAGCAGATGTCTGAT 134

Qy   1765 aagcaactgctggccacctggaaaaggataatccaggcacatggcggcactgttgacccc 1824
Db    133 AAGCAACTGCTGGCCACCTGGAAAAGGATAATCCAGGCTCATGGCGGCACTGTTGACCCC 74

Qy   1825 accttcacgagtcgatgcacgcaccttctctgtgagagtcaagtcagcagcgcgtatgca 1884
Db     73 ACCTTCACGAGTCGATGCACGCACCTTCTCTGTGAGAGTCAAGTCAGCAGCGCGTATGCA 14

Qy   1885 caggcaataaagag 1897
Db     13 CAGGCAATAAGAG 1
```

RESULT 13
G15998/c
LOCUS G15998 687 bp DNA linear STS 19-JAN-1996
DEFINITION human STS CHLC.GCT8H06.P11207 clone GCT8H06, sequence tagged site.
ACCESSION G15998
VERSION G15998.1 GI:1161887
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE human vector=pJCP1 host=E.coli dut+ung+ (DH10B) Marker Selected genomic DNA prepared from XY individual of French nationality.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 687)
AUTHORS Murray,J., Sheffield,V, Weber,J.L., Duyk,G. and Buetow,K.H.
TITLE Cooperative Human Linkage Center
JOURNAL Unpublished
COMMENT Synonyms: GCT8H06, CHLC.GCT8H06.T11206
Contact: Dr. Jeffrey C. Murray
Uofi
The University of Iowa
Department of Pediatrics, Iowa City, IA 52242, USA
Tel: (319) 356-3508
Fax: (319) 356-3347
Email: jeff-murray@uiowa.edu

Primer A: CCCCACACTCAAACACTCTC
Primer B: GACACACATGCTACAGCAGC
STS size: 347
PCR Profile:
denature: 30 seconds at 94 degrees C
annealing: 75 seconds at 55 degrees C
extension: 15 seconds at 72 degrees C
PCR cycles: 27
extension: 6 minutes at 72 degress C
Protocol:
Template: 30ng genomic DNA

```
                     Primer:          each 1.5 pmole
                     dNTPs:           each 200 uM
                     Taq Polymerase:  0.3 units
                     Total Vol:       10 ul

          Buffer:
                     MgCl2:  1.5mM
                     KCl:     50mM
                     Tris:    10mM
                     pH:     8.3.
FEATURES             Location/Qualifiers
     source          1. .687
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     STS             127. .473
     primer_bind     127. .146
     primer_bind     complement(454. .473)
BASE COUNT      167 a    148 c    164 g    203 t      5 others
ORIGIN
```

Query Match 7.4%; Score 263.4; DB 11; Length 687;
Best Local Similarity 79.9%; Pred. No. 5.1e-47;
Matches 345; Conservative 0; Mismatches 56; Indels 31; Gaps 2;

```
Qy   1318 atgaaccggccatcaaatgtagcacatatcttacagactctttcagcacctacgaaaaat 1377
          |||   |  | | |   |||     | ||||||||||||||||||| ||||||||||||
Db    616 ATGTTTCACCTAACTTTTGTTTTGTAGATCTTACAGACTCTTTCAGNACCTACGAAAAAT 557

Qy   1378 ttagaacagcaggtgaatcacagccagcagggacatacaaatgccaatgcagtgctgttt 1437
          ||||  ||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    556 TTAGNGCAGNAGGTCAATCACAGCCAGCAGGGACATACAAATGCCAATGCAGTGCTGTTT 497

Qy   1438 agccaagtgaaagtgactccagagacacacatgctacagcagcagcagcaggcccagcag 1497
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    496 AGCCAAGTGAAAGTGACTCCAGAGACACACATGCTACAGCAGCAGCAGCAGGCCCAGCAG 437

Qy   1498 cagcagcagcagcacccggttttacaccttcagccccagcagataatgcagctccagcag 1557
          |||||||||||||||||||||||||||||||||||||||||| || ||||||||||||||
Db    436 CAGCAGCAGCAGCACCCGGTTTTACACCTTCAGCCCCAGCAGNTANTGCAGCTCCAGCAG 377

Qy   1558 cagcagcagcagcaga----------------------tctctcagcaaccttacccc 1593
          ||||||||||||||||                      |   ||||    |  |  |
Db    376 CAGCAGCAGCAGCAGATGCAAAGTCAGACAGCGCCACACTTGAGTCAGACGTCACAGGCG 317

Qy   1594 cagcagccgccgcatccattttca-------cagcagcagcagcagcaacagcagccacc 1646
          | |||||  | |  ||||   |||       | ||||||||||||||||||||||||||
Db    316 CTGCAGCATCAGGTTCCACCCTCAGCAGCCCCCGCAGCAGCAGCAGCAACAGCAGCCACC 257

Qy   1647 accatcgcctcagcagcatcagctttttggacatgatccagcagtggagattccagaaga 1706
          ||||||||||||||||||||||||||||||||||||||||||||||||||| |    ||
Db    256 ACCATCGCCTCAGCAGCATCAGCTTTTTGGACATGATCCAGCAGTGGAGAGTTAGTAATG 197

Qy   1707 aggcttcttatt 1718
            | || || ||
Db    196 GAGATTATTTTT 185
```

```
RESULT  14
AX186240
LOCUS       AX186240                690 bp    DNA     linear   PAT 06-AUG-2001
DEFINITION  Sequence 1935 from Patent WO0142467.
ACCESSION   AX186240
VERSION     AX186240.1  GI:15137668
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 690)
  AUTHORS   Schlegel,R., Deeds,J., Berger,A. and Zhao,X.
  TITLE     Genes, compositions, kits, and methods for identification,
            assessment, prevention, and therapy of cervical cancer
  JOURNAL   Patent: WO 0142467-A 1935 14-JUN-2001;
            Millennium Predictive Medicine, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .690
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      218 a    174 c    172 g    124 t      2 others
ORIGIN
```

Query Match 5.7%; Score 204.6; DB 6; Length 690;
Best Local Similarity 95.9%; Pred. No. 3.7e-34;
Matches 210; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```
Qy   1073 cggccgaagtcccacagttagctgcagcaaaacgcaggctgcctcagggaaaggagcctg 1132
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     23 CGGCCGAAGTCCCACAGTTAGCTGCAGCAAAACGCAGGCTGCCTCAGGGAAAGGAGCCTG 82

Qy   1133 ggttgattaacttgtgtgccaatgtcccacccgtcccaggtaacattttgccccctgagg 1192
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     83 GGTTGATTAACTTGTGTGTCAATGTCCCACCCGTCCCAGGTAACATTTTGCCCCCTGAGG 142

Qy   1193 tccggggtaatttaatggctgctggacaaaacctccaaagttctgaaagatcagaaatga 1252
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    143 TCCGGGGTAATTTAATGGCTGCTGGACAAAACCTCCAAAGTTCTGAAAGATCAGAAATGA 202

Qy   1253 tagctacctggagtccagctgtacggacactgaggaata 1291
          |||||||||||||||||||||||||||    ||  | | |
Db    203 TAGCTACCTGGAGTCCAGCTGTACGGCACTTGGCGTAAA 241
```

```
RESULT  15
AX187445
LOCUS       AX187445                504 bp    DNA     linear   PAT 06-AUG-2001
DEFINITION  Sequence 3140 from Patent WO0142467.
ACCESSION   AX187445
VERSION     AX187445.1  GI:15138906
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 504)
  AUTHORS   Schlegel,R., Deeds,J., Berger,A. and Zhao,X.
  TITLE     Genes, compositions, kits, and methods for identification,
            assessment, prevention, and therapy of cervical cancer
  JOURNAL   Patent: WO 0142467-A 3140 14-JUN-2001;
            Millennium Predictive Medicine, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .504
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      147 a    135 c    121 g    101 t
ORIGIN
```

Query Match 5.2%; Score 186.8; DB 6; Length 504;
Best Local Similarity 94.0%; Pred. No. 2.8e-30;
Matches 205; Conservative 0; Mismatches 12; Indels 1; Gaps 1;

```
Qy   1075 gccgaagtcccacagttagctgcagcaaaacgcaggctgcctcagggaaaggagcctggg 1134
          | |  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 GGCCGAGTCCCACAGTTAGCTGCAGCAAAACGCAGGCTGCCTCAGGGAAAGGAGCCTGGG 80

Qy   1135 ttgattaacttgtgtgccaatgtcccacccgtcccaggtaacattttgccccctgaggtc 1194
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     81 TTGATTAACTTGTGTGTCAATGTCCCACCCGTCCCAGGTAACATTTTGCCCCCTGAGGTC 140

Qy   1195 cggggtaatttaatggctgctggacaaaacctccaaagttc-tgaaagatcagaaatgat 1253
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    141 CGGGGTAATTTAATGGCTGCTGGACAAAACCTCCAAAGTTCTTGAAAGATCAGAAATGAT 200

Qy   1254 agctacctggagtccagctgtacggacactgaggaata 1291
```

```
                    ||||||||||||||||||||||||||    ||  | | |
Db    201 AGCTACCTGGAGTCCAGCTGTACGGCACTTGGCGTAAA 238
```

Search completed: June 4, 2002, 15:53:06
Job time: 17024 sec

EST5

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 4, 2002, 11:09:22 ; Search time 3648.51 Seconds
(without alignments)
13243.516 Million cell updates/sec

Title: US-09-664-641-10
Perfect score: 3580
Sequence: 1 catagacagtcgatacgcct.....gaacttagaaaaaaaaaaaa 3580

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		809.4	22.6	1059	10	BM470958	BM470958 AGENCOURT
c 2		707.2	19.8	760	9	AI671887	AI671887 wb41c03.x
c 3		700.6	19.6	780	9	AW073165	AW073165 wy94g07.x
c 4		698.2	19.5	811	9	AI357401	AI357401 qy13d06.x
5		653.2	18.2	759	10	BE260052	BE260052 601150905
c 6		650.4	18.2	652	9	AI133662	AI133662 HA2284 Hu
c 7		639.8	17.9	680	9	AI741935	AI741935 wg29g05.x
8		583	16.3	669	10	BG900296	BG900296 HOA37-1-B
9		581.2	16.2	625	9	AU123803	AU123803 AU123803
10		573.2	16.0	803	10	BE541477	BE541477 601067929
c 11		558.4	15.6	597	9	AW009159	AW009159 ws78b09.x
c 12		547.2	15.3	552	9	AW004836	AW004836 ws93c05.x
c 13		540.4	15.1	542	10	BF939756	BF939756 nac83a11.
c 14		539.6	15.1	598	9	AI807116	AI807116 we85g11.x
c 15		528.6	14.8	551	9	AI393049	AI393049 tg25c09.x
c 16		526.8	14.7	541	9	AA846820	AA846820 aj99g08.s
c 17		518.8	14.5	535	9	AW937777	AW937777 QV3-DT004
18		507.2	14.2	1154	10	BM461168	BM461168 AGENCOURT
19		500.8	14.0	504	10	BF876036	BF876036 IL3-ET011
20		499	13.9	516	9	AW390801	AW390801 RC6-ST019
21		499	13.9	531	9	AL135690	AL135690 DKFZp762P
22		496	13.9	730	10	BI248290	BI248290 602966883
23		485.8	13.6	677	10	BG259569	BG259569 602378702
c 24		473	13.2	514	9	AI625826	AI625826 ty65b06.x
c 25		472	13.2	472	9	AI362238	AI362238 qy45b09.x
c 26		471.8	13.2	520	9	AW341490	AW341490 hd10d04.x
c 27		467.8	13.1	504	9	AA587951	AA587951 nm88g02.s
c 28		466.8	13.0	480	9	AW303585	AW303585 xv20h08.x
c 29		463.8	13.0	540	10	BF916607	BF916607 IL3-UT011
c 30		454.4	12.7	480	10	N51347	N51347 yz15g08.s1
c 31		445	12.4	445	10	BE677288	BE677288 7d83a09.x
32		441.2	12.3	457	9	AL042281	AL042281 DKFZp434L
33		439.4	12.3	888	10	BE573124	BE573124 601331567
34		438	12.2	872	10	BI735116	BI735116 603356256
c 35		434.6	12.1	640	10	BI401433	BI401433 MI-P-CP0-
c 36		433.4	12.1	435	9	AI080755	AI080755 ox54h03.x
37		432.4	12.1	547	9	AW867972	AW867972 RC0-SN004
c 38		425.2	11.9	430	9	AW117548	AW117548 xd93b10.x
39		424.4	11.9	426	9	AW663176	AW663176 hh75f01.y
c 40		416.4	11.6	419	9	AW183769	AW183769 xj87f05.x
41		416.4	11.6	583	10	BF320169	BF320169 uz51c09.y
42		412.6	11.5	738	10	BI684354	BI684354 603308191
c 43		412.4	11.5	425	9	AA917929	AA917929 ol67q01.s
c 44		411.2	11.5	506	10	N51997	N51997 yz07h04.s1
45		411.2	11.5	555	10	BF012553	BF012553 ux57g08.y

ALIGNMENTS

RESULT 1
BM470958
LOCUS BM470958 1059 bp mRNA linear EST 05-FEB-2002
DEFINITION AGENCOURT_6462994 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5533689 5', mRNA sequence.
ACCESSION BM470958
VERSION BM470958.1 GI:18520000
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1059)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12218 row: p column: 10
High quality sequence stop: 751.
FEATURES Location/Qualifiers
source 1. .1059
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5533689"
/clone_lib="NIH_MGC_71"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.1 kb. "
BASE COUNT 307 a 245 c 242 g 260 t 5 others
ORIGIN

```
  Query Match            22.6%;  Score 809.4;  DB 10;  Length 1059;
  Best Local Similarity  97.3%;  Pred. No. 4.1e-124;
  Matches  896;  Conservative    0;  Mismatches   18;  Indels    7;  Gaps    7;

Qy   1747 ccagaqcaqatgtctgataagcaactgctggccacctggaaaaggataatccaggcacat 1806
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CTAGAGCAGATGTCTGATAAGCAACTGCTGGCCACCTGGAAAAGGATAATCCAGGCACAT 60

Qy   1807 ggcggcactgttgaccccaccttcacgagtcgatgcacgcaccttctctgtgagagtcaa 1866
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GGCGGCACTGTTGACCCCACCTTCACGAGTCGATGCACGCACCTTCTCTGTGAGAGTCAA 120

Qy   1867 gtcagcagcgcgtatgcacaggcaataagagaaagaaaagagatgtgttactgcacactgg 1926
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GTCAGCAGCGCGTATGCACAGGCAATAAGAGAAAGAAACAGATGTGTTACTGCACACTGG 180

Qy   1927 ttaaacacagtcttaaaaaaaaaaaaaatggtaccgccgcaccgagcccttcacttccca 1986
          |||||||||||||||| || || ||||||||||||||||||||||||||||||||||||
Db    181 TTAAACACAGTCTTAAAGAAGAAGAAAATGGTACCGCCGCACCGAGCCCTTCACTTCCCA 240

Qy   1987 gtggccttcccaccaggaggaaagccatgttcacagcatattatttctgtgactggattt 2046
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GTGGCCTTCCCACCAGGAGGAAAGCCATGTTCACAGCATATTATTTCTGTGACTGGATTT 300

Qy   2047 gttgatagtgacagagatgacctaaaattaatggcttatttggcaggtgccaaatatacg 2106
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GTTGATAGTGACAGAGATGACCTAAAATTAATGGCTTATTTGGCAGGTGCCAAATATACG 360

Qy   2107 ggttatctatgccgcagcaacacagtcctcatctgtaaagaaccaactggtttaaagtat 2166
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GGTTATCTATGCCGCAGCAACACAGTCCTCATCTGTAAAGAACCAACTGGTTTAAAGTAT 420

Qy   2167 gaaaaagccaaagagtggaggatacccctgtgtcaacgcccagtggcttggcgacattctt 2226
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GAAAAAGCCAAAGAGTGGAGGATACCCTGTGTCAACGCCCAGTGGCTTGGCGACATTCTT 480

Qy   2227 ctgggaaactttgaggcactgaggcagattcagtatagtcgctacacggcattcagtctg 2286
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CTGGGAAACTTTGAGGCACTGAGGCAGATTCAGTATAGTCGCTACACGGCATTCAGTCTG 540

Qy   2287 caggatccatttgcccctacccagcatttagttttaaatctttagatgcttggagagtt 2346
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CAGGATCCATTTGCCCCTACCCAGCATTTAGTTTTAAATCTTTTAGATGCTTGGAGAGTT 600

Qy   2347 cccttaaaagtgtctgcagagttgttgatgagtataagactacctcccaaactgaaacag 2406
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CCCTTAAAAGTGTCTGCAGAGTTGTTGATGAGTATAAGACTACCTCCCAAACTGAAACAG 660

Qy   2407 aatgaagtagctaatgtccagccttcttcccaaagagccagaattgaagacgtaccacct 2466
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    661 AATGAAGTAGCTAATGTCCAGCCTTCTTCCAAAAGAGCCAGAATTGAAGACGTACCACCT 720

Qy   2467 cccactaaaaagctaactccagaattga-ccccttttgtgcttttcactggattcgagcc 2525
          |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db    721 CCCACTAAAAAGCTAACTCCAGAATTGACCCCCTTTTGTGCTTTTCACTGGATTCGAGCC 780

Qy   2526 tgtccaggttcaacagtatattaagaagctctacattcttg-gtggagaggttgcg-gag 2583
          ||||||||    |||||||||||||||||||||||||||| |||||||||||||| ||| 
Db    781 TGTCCAGGNNTCACAGTATATTAAGAAGCTCTACATTCTTGNGTGGAGAGGTTGCGNGAG 840

Qy   2584 tctgcac-agaagtgcacacacctca-ttgccagcaaagt-gactcgca-ccgtgaagtt 2639
          ||||||| |||||||||||||||||| |||||||||||| ||||||| ||||||||||||
Db    841 TCTGCACAAGAAGTGCACACACCTCATTTGCCAGCAAAGTGGACTCGCACCCGTGAAGTT 900

Qy   2640 cctgacggcgatttctgtcgt 2660
          ||||   | |  || | | ||
Db    901 CCTGGACGGGGATTTTCTGGT 921


RESULT   2
AI671887/c
```

```
LOCUS       AI671887                760 bp    mRNA    linear   EST 17-DEC-1999
DEFINITION  wb41c03.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2308228 3'
            similar to TR:O15404 O15404 CAGF28 ;contains LTR5.b2 LTR5
            repetitive element ;, mRNA sequence.
ACCESSION   AI671887
VERSION     AI671887.1  GI:4851618
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 760)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1444    Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 458.
FEATURES             Location/Qualifiers
     source          1. .760
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2308228"
                     /clone_lib="NCI_CGAP_GC6"
                     /tissue_type="pooled germ cell tumors"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA
                     from the normalized library NCI_CGAP_GC4 was prepared, and
                     ss circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from a pool
                     of 5,000 clones made from the same library (cloneIDs
                     1257096-1258631, 1469064-1470983, and 1475592-1476743).
                     Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      260 a    153 c    133 g    210 t      4 others
ORIGIN
```

```
  Query Match            19.8%;  Score 707.2;  DB 9;  Length 760;
  Best Local Similarity  96.3%;  Pred. No. 3.4e-107;
  Matches  732;  Conservative    0;  Mismatches   27;  Indels    1;  Gaps    1;

Qy   2810 cactctttaaggcaaaatatttttacatcacacctggaatctgcccaagtctttccacta 2869
          || | | |   | |||  |||||||||||||  ||| |||| |||||||||| |||||
Db    759 CATTTTATTTAGGCAAAATATTTTACATCACACTGGGATCCTGCNCAAGTCTTTACACTA 700

Qy   2870 tgaaggcaatcgtagagtgtgcaggaggaaaggtgttatccaagcagccatctttccgga 2929
          ||||||| ||||||||||||   |||| ||||||||   ||||||||||||||||||| |
Db    699 TGAAGGCCATCGTAGAGTGTCAGGGAGNAAAGGTGTATNCCAAGCAGCCATCTTTCCGNA 640

Qy   2930 agctcatggagcacaagcagaactcgagtttgtcggaaataatttttaatatcctgtgaaa 2989
          |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db    639 AGCTCATGGAGCAC-AGCAGAACTCGAGTTTGTCGGAAATAATTTTTAATATCCTGTGAAA 581

Qy   2990 atgaccttcatttatgccgagaatattttgccagaggcatagatgttcacaatgcagagt 3049
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    580 ATGACCTTCATTTATGCCGAGAATATTTTGCCAGAGGCATAGATGTTCACAATGCAGAGT 521

Qy   3050 tcgttctgactggagtgctcactcaaacgctggactatgaatcatataagtttaactgat 3109
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    520 TCGTTCTGACTGGAGTGCTCACTCAAACGCTGGACTATGAATCATATAAGTTTAACTGAT 461

Qy   3110 ggcgtctaggctgccgtgcatgtcgactcctgcggtgcggggctggctgtctggctggcg 3169
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    460 GGCGTCTAGGCTGCCGTGCATGTCGACTCCTGCGGTGCGGGGCTGGCTGTCTGGCTGGCG 401

Qy   3170 aggagctgctgcgcttccttcacatgctcttgttttccagctgctttcctgggggatcag 3229
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    400 AGGAGCTGCTGCGCTTCCTTCACATGCTCTTGTTTTCCAGCTGCTTTCCTGGGGGATCAG 341

Qy   3230 actgtgaagcaggaagacagatataataaatatactgcatctttttaagatgtgcaatttt 3289
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    340 ACTGTGAAGCAGGAAGACAGATATAATAAATATACTGCATCTTTTTAAGATGTGCAATTT 281

Qy   3290 tattctgaggaaacataaattatgttttgtattatatgact taagagcccacattaggt 3349
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    280 TATTCTGAGGAAACATAAATTATGTTTTGTATTATATGACTTTAAGAGCCCACATTAGGT 221

Qy   3350 tttatgattcatttgccaggtttttaaatgtttcacaaaactgttacgggacttcaact 3409
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    220 TTTATGATTCATTTGCCAGGTTTTTAAATGTTTCACAAAACTGTTACGGGACTTCAACT 161

Qy   3410 agaaataaaatggtgtaaataaagaccttgctatctctaaattatggatgttaaagattt 3469
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    160 AGAAATAAAATGGTGTAAATAAAGACCTTGCTATCTCTAAATTATGGATGTTAAAGATTT 101

Qy   3470 gaaatgttttgtactttgattatttttattcctatactctgttttcttttatattgata 3529
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    100 GAAATGTTTTGTACTTTGATTATTTTTATTCCTATACTCTGTTTTCTTTTATATTGATA 41

Qy   3530 tcttgcccacatttaaaataaatgtacttttgaacttaga 3569
          ||||||||||||||||||||||||||||||||||||| |
Db     40 TCTTGCCCACATTTTAAATAAATGTACTTTTGAACTTAAA 1


RESULT   3
AW073165/c
LOCUS       AW073165                780 bp    mRNA    linear   EST 09-MAR-2000
DEFINITION  wy94g07.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:2556252 3'
            similar to TR:O15404 O15404 CAGF28 ;contains LTR5.b2 LTR5
            repetitive element ;, mRNA sequence.
ACCESSION   AW073165
VERSION     AW073165.1  GI:6028163
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 780)
  AUTHORS   NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGAP), Tumor Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Robert SLrausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
            Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 3034   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 455.
FEATURES             Location/Qualifiers
     source          1. .780
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2556252"
                     /clone_lib="NCI_CGAP_Brn23"
                     /tissue_type="glioblastoma (pooled)"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATATCTTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library is normalized, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT      254 a    157 c    154 g    213 t      2 others
ORIGIN


  Query Match           19.6%;  Score 700.6;  DB 9;  Length 780;
  Best Local Similarity  96.6%;  Pred. No. 4.1e-106;
  Matches  758;  Conservative    0;  Mismatches   21;  Indels    6;  Gaps    4;

Qy   2785 tccttaaaacgggcacacgtttctccactctttaaggcaaaatatttttacatcacacct 2844
          ||| | ||| |||||||||||     || | ||  || |  |||   || |||||| |||||
Db    780 TCCCTTAAAAGGGCACACGT---CCCNCCCTCNAAAGGCAAAATATTCTACATC-CACCT 725

Qy   2845 ggaatctgcccaagtctttccactatgaaggcaatcgtagagtgtgcaggaggaaaggtg 2904
          |||||||||||||||||| |||||||||||||||||||||||||||||| ||||||||||
Db    724 GGAATCTGCCCAAGTCTTCCCACTATGAAGGCAATCGTAGAGTGTGCAGCAGGAAAGGTG 665

Qy   2905 ttatccaagcagccatctttccggaagctcatggagcacaagcagaactcgagtttgtcg 2964
          |||||||||||||||||| ||| ||||||||||||||||||||||||||||||||||||
Db    664 -TATCCAAGCAGCCATCTTCCCGCAAGCTCATGGAGCACAAGCAGAACTCGAGTTTGTCG 606

Qy   2965 gaaataatttttaatatcctgtgaaaatgaccttcatttatgccgagaatattttgccaga 3024
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    605 GAAATAATTTTTAATATCCTGTGAAAATGACCTTCATTTATGCCGAGAATATTTTGCCAGA 546

Qy   3025 ggcatagttgttcacaatgcagagttcgttctgactggagtgctcactcaaacgctggac 3084
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    545 GGCATAGATGTTCACAATGCAGAGTTCGTTCTGACTGGAGTGCTCACTCAAACGCTGGAC 486

Qy   3085 tatgaatcatataagtttaactgatggcgtctaggctgccgtgcatgtcgactcctgcgg 3144
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    485 TATGAATCATATAAGTTTAACTGATGGCGTCTAGGCTGCCGTGCATGTCGACTCCTGCGG 426

Qy   3145 tgcggggctggctgtctggctggcgaggagctgctgcgcttccttcacatgctcttgttt 3204
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    425 TGCGGGGCTGGCTGTCTGGCTGGCGAGGAGCTGCTGCGCTTCCTTCACATGCTCTTGTTT 366

Qy   3205 tccagctgctttcctgggggatcagactgtgaagcaggaagacagatataataaatatac 3264
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    365 TCCAGCTGCTTTCCTGGGGGATCAGACTGTGAAGCAGGAAGACAGATATAATAAATATAC 306

Qy   3265 tgcatctttttaagatgtgcaattttattctgaggaaacataaattatgttttgtattat 3324
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    305 TGCATCTTTTTAAGATGTGCAATTTTATTCTGAGGAAACATAAATTATGTTTTGTATTAT 246

Qy   3325 atgactttaagag-cccacattaggttttatgattcatttgccaggtttttaaatgtttt 3383
          |||||||||||||| ||| ||||||||||||||||||||||||||||||||||||||||
Db    245 ATGACTTTAAGAGCCCCCCATTAGGTTTTATGATTCATTTGCCAGGTTTTTAAATGTTTT 186

Qy   3384 cacacaactgttacgggacttcaactagaaataaaatggtgtaaataaagaccttgctat 3443
          |||| |||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db    185 CACAAAACTGTTACGGGACTTCAACTAGAAATAAAATGGTGTAAATAAAGCCCTTGCTAT 126

Qy   3444 ctctaaattatggatgttaaagatttgaaatgttttgtactttgattatttttatttctt 3503
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    125 CTCTAAATTATGGATGTTAAAGATTTGAAATGTTTTGTACTTTGATTATTTTTATTTCTT 66

Qy   3504 atactctgttttcttttatattgatatcttgcccacattttaaataaatgtacttttgaa 3563
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     65 ATACTCTGTTTTCTTTTATATTGATATCTTGCCCACATTTTAAATAAATGTACTTTTGAA 6

Qy   3564 cttag 3568
```

```
          |||||
Db      5 CTTAG 1


RESULT   4
AI357401/c
LOCUS       AI357401                811 bp    mRNA    linear   EST 15-FEB-1999
DEFINITION  qy13d06.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:2011883 3'
            similar to TR:O15404 O15404 CAGF28 ;, mRNA sequence.
ACCESSION   AI357401
VERSION     AI357401.1  GI:4109022
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 811)
  AUTHORS   NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGAP), Tumor Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
            Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1784   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 364.
FEATURES             Location/Qualifiers
     source          1. .811
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2011883"
                     /clone_lib-"NCI_CGAP_Brn23"
                     /tissue_type-"glioblastoma (pooled)"
                     /lab_host-"DH10B"
                     /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATATCTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library is normalized, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT      271 a    163 c    147 g    230 t
ORIGIN


  Query Match           19.5%;  Score 698.2;  DB 9;  Length 811;
  Best Local Similarity  94.2%;  Pred. No. 1e-105;
  Matches  768;  Conservative    0;  Mismatches   43;  Indels    4;  Gaps    4;

Qy   2763 tttctctttcagcttggaagaatcctaaaacgggcacacgtttctccactctttaaggc 2822
           ||| | |||||||| |  | |||||| ||| |||||| ||| ||| |||||||||| |
Db    811 TTTTTTTTTCAGCTGGAGAAAATCCTAAAACGGGGCAC-GGTTCTTCCCCTCTTTAAGCC 753

Qy   2823 aaaatatttttacatcacacctggaatctgcccaagtctttccactatgaaggcaatcgt 2882
           |||||| |||| |||||||| |||| || |  |||||||| | | || ||||
Db    752 AAAATAATTTT-CATCACACTGGGAAATCTGCCCAAGTCTTCCACTATAAGGCCATTCGT 694

Qy   2883 agagtgtgcaggaggaaaggtgttatccaagcagccatctttccggaagctcatggagca 2942
           | || | ||||  | |||||||||||| ||| | || |||| | ||||||||||||||||
Db    693 A-AGGGCGCAGAGGAAAAGGTGTTATCAAAG-ACCCTTCTTCCGGGAAGCTCATGGAGCA 636

Qy   2943 caagcagaactcgagtttgtcggaaataatttaatatcctgtgaaaatgaccttcattt 3002
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     635 CAAGCAGAACTCGAGTTTGTCGGAAATAATTTTAATATCCTGTGAAAATGACCTTCATTT 576

Qy   3003 atgccgagaatattttgccagaggcatagatgttcacaatgcagagttcgttctgactgg 3062
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     575 ATGCCGAGAATATTTTGCCAGAGGCATAGATGTTCACAATGCAGAGTTCGTTCTGACTGG 516

Qy   3063 agtgctcactcaaacgctggactatgaatcatataagtttaactgatggcgtctaggctg 3122
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     515 AGTGCTCACTCAAACGCTGGACTATGAATCATATAAGTTTAACTGATGGCGTCTAGGCTG 456

Qy   3123 ccgtgcatgtcgactcctgcggtgcggggctggctgtctggctggcgaggagctgctgcg 3182
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     455 CCGTGCATGTCGACTCCTGCGGTGCGGGGCTGGCTGTCTGGCTGGCGAGGAGCTGCTGCG 396

Qy   3183 cttccttcacatgctcttgttttccagctgctttcctgggggatcagactgtgaagcagg 3242
           |||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     395 CTTCTTCCACATGCTCTTGTTTTCCAGCTGCTTTCCTGGGGGATCAGACTGTGAAGCAGG 336

Qy   3243 aagacagatataataaatatactgcatctttttaagatgtgcaatttttattctgaggaaa 3302
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     335 AAGACAGATATAATAAATATACTGCATCTTTTTAAGATGTGCAATTTTATTCTGAGGAAA 276

Qy   3303 cataaattatgttttgtattatatgactttaagagcccacattaggttttatgattcatt 3362
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     275 CATAAATTATGTTTTGTATTATATGACTTTAAGAGCCCACATTAGGTTTTATGATTCATT 216

Qy   3363 tgccaggtttttaaatgttttcacaaaactgttacgggacttcaactagaaataaaatgg 3422
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     215 TGCCAGGTTTTTAAATGTTTTCACAAAACTGTTACGGGACTTCAACTAGAAATAAAATGG 156

Qy   3423 tgtaaataaagaccttgctatctctaaattatggatgttaaagatttgaaatgttttgta 3482
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     155 TGTAAATAAAGACCTTGCTATCTCTAAATTATGGATGTTAAAGATTTGAAATGTTTTGTA 96

Qy   3483 ctttgattatttttatttcttatactctgttttctttttatattgatatcttgcccacatt 3542
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      95 CTTTGATTATTTTTATTTCTTATACTCTGTTTTCTTTTATATTGATATCTTGCCCACATT 36

Qy   3543 ttaaataaatgtacttttgaacttagaaaaaaaaa 3577
           |||||||||||||||||||||||||||||||||
Db      35 TTAAATAAATGTACTTTTGAACTTAGAAAAAAAAA 1


RESULT   5
BE260052
LOCUS       BE260052                759 bp    mRNA    linear   EST 26-OCT-2000
DEFINITION  601150905F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3503689 5',
            mRNA sequence.
ACCESSION   BE260052
VERSION     BE260052.1  GI:9131032
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 759)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM177  row: i  column: 02
            High quality sequence stop: 731.
FEATURES             Location/Qualifiers
     source          1. .759
```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3503689"
/clone_lib="NIH_MGC_19"
/tissue_type="neuroblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 212 a 179 c 177 g 191 t
ORIGIN

Query Match 18.2%; Score 653.2; DB 10; Length 759;
Best Local Similarity 96.0%; Pred. No. 2.8e-98;
Matches 724; Conservative 0; Mismatches 23; Indels 7; Gaps 5;

```
Qy   1961 cgccgcaccgaqccctcacttcccaqtqgcctcccaccaggaggaaagccatgttcac 2020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      9 CGCCGCACCGAGCCCTTCACTTCCCAGTGGCCTTCCCACCAGGAGGAAAGCCATGTTCAC 68

Qy   2021 agcatattatttctgtgactggatttgttgatagtgacagagatgacctaaaattaatgg 2080
          |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db     69 AGCATATTATTTCTGTGACTGGA-TTGTTGATAGTGACAGAGATGACCTAAAATTAATGG 127

Qy   2081 cttatttggcaggtgccaaatatacgggttatctatgccgcagcaacacagtcctcatct 2140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    128 CTTATTTGGCAGGTGCCAAATATACGGGTTATCTATGCCGCAGCAACACAGTCCTCATCT 187

Qy   2141 qtaaagaaccaactggtttaaagtatgaaaaagccaaagagtggaggataccctgtgtca 2200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    188 GTAAAGAACCAACTGGTTTAAAGTATGAAAAAGCCAAAGAGTGGAGGATACCCTGTGTCA 247

Qy   2201 acgcccaqtggcttggcgacattcttctgggaaactttgaggcactgaggcagattcagt 2260
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    248 ACGCCCAGTGGCTTGGCGACATTCTTCTGGGAAAC-TTGAGGCACTGAGGCAGATTCAGT 306

Qy   2261 atagtcgctacacggcattcagtctgcaggatccatttgcccctacccagcatttagttt 2320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    307 ATAGTCGCTACACGGCATTCAGTCTGCAGGATCCATTTGCCCCTACCCAGCATTTAGTTT 366

Qy   2321 taaatcttttagatgcttggagagttcccttaaaagtgtctgcagagttgttgatgagta 2380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    367 TAAATCTTTTAGATGCTTGGAGAGTTCCCTTAAAAGTGTCTGCAGAGTTGTTGATGAGTA 426

Qy   2381 taagactacctcccaaactgaaacagaatgaagtagctaatgtccagccttcttcccaaa 2440
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    427 TAAGACTACCTCCCAAACTGAAACAGAATGAAGTAGCTAATGTCCAGCCTTCTTCCAAAA 486

Qy   2441 gagccagaattgaagacgtaccacctcccactaaaaagctaactccagaattgacccctt 2500
          |||||||||||||||||||||||||||||||||||||||||||||||| |||||| | |
Db    487 GAGCCAGAATTGAAGACGTACCACCTCCCACTAAAAAGCTAACTCCAGAATGGACCCC-T 545

Qy   2501 ttgtgcttttcactggattcgagcctgtccaggttcaacagtatattaagaagctctaca 2560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    546 TTGTGCTTTTCACTGGATTCGAGCCTGTCCAGGTTCAACAGTATATTAAGAAGCTCTACA 605

Qy   2561 ttcttggtggagaggttgcggagtctgcacagaagtgcacacacctcattgccagcaaag 2620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    606 TTCTTGGTGGAGAGGTTGCGGAGTCTGCACAGAAGTGCACACACCTCATTGCCAGCAAAG 665

Qy   2621 tgactcg-caccgtgaagttcctgacggcgatt---tctgtcgtgaagcacatagtgacg 2676
          ||||||| |||||  ||||||||||||||||    | |      ||||||||||||||||
Db    666 TGACTCGCCACCGGAAAGTTCCTGACGGCGATTTTCTGTCGTTGAAAGCACATAGTGACG 725

Qy   2677 ccagagtggctggaagaatgcttcaggtgtcaga 2710
          ||     |   | |||||||  || |||||||||
Db    726 CCGAGTGGCTGGAAAGAATGGCTCCGGTGTCAGA 759
```

RESULT 6
AI133662/c
LOCUS AI133662 652 bp mRNA linear EST 11-NOV-1999
DEFINITION HA2284 Human fetal liver cDNA library Homo sapiens cDNA, mRNA sequence.
ACCESSION AI133662
VERSION AI133662.1 GI:6360978
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 652)
AUTHORS Yu,Y., Zhang,C., Luo,L., Ouyang,S., Li,W., Wu,J., Zhou,S., Liu,M. and He,F.
TITLE Expression profile analysis of a human fetal liver cDNA library
JOURNAL Unpublished (1998)
COMMENT Contact: Yongtao Yu
Department of Hematology
Beijing Institute of Radiation Medicine
27 Taiping Road, Beijing 100850, P.R.China
Tel: 0086-10-68159479
Fax: 0086-10-68214653
Email: yyt48@yahoo.com.
FEATURES Location/Qualifiers
source 1. .652
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Human fetal liver cDNA library"
/tissue_type="liver"
/dev_stage="fetal"
/lab_host="MC1061/P3"
/note="Vector: pCDNA1"
BASE COUNT 175 a 166 c 147 g 164 t
ORIGIN

Query Match 18.2%; Score 650.4; DB 9; Length 652;
Best Local Similarity 99.8%; Pred. No. 8.5e-98;
Matches 651; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy   2558 acattcttggtggagaggttgcggagtctgcacagaagtgcacacacctcattgccagca 2617
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     652 ACATTCTTGGTGGAGAGGTTGCGGAGTCTGCACAGAAGTGCACACACCTCATTGCCAGCA 593

Qy   2618 aagtgactcgcaccgtgaagttcctgacggcgatttctgtcgtgaagcacatagtgacgc 2677
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     592 AAGTGACTCGCACCGTGAAGTTCCTGACGGCGATTTCTGTCGTGAAGCACATAGTGACGC 533

Qy   2678 cagagtggctggaagaatgcttcaggtgtcagaagttcattgatgagcagaactacattc 2737
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     532 CAGAGTGGCTGGAAGAATGCTTCAGGTGTCAGAAGTTCATTGATGAGCAGAACTACATTC 473

Qy   2738 tccgagatgctgaggcagaagtacttttctctttcagcttggaagaatccttaaaacggg 2797
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     472 TCCGAGATGCTGAGGCAGAAGTACTTTTCTCTTTCAGCTTGGAAGAATCCTTAAAACGGG 413

Qy   2798 cacacgtttctccactctttaaggcaaaatatttttacatcacacctggaatctgcccaa 2857
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     412 CACACGTTTCTCCACTCTTTAAGGCAAAATATTTTTACATCACACCTGGAATCTGCCCAA 353

Qy   2858 gtctttccactatgaaggcaatcgtagagtgtgcaggaggaaaggtgttatccaagcagc 2917
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     352 GTCTTTCCACTATGAAGGCAATCGTAGAGTGGGCAGGAGGAAAGGTGTTATCCAAGCAGC 293

Qy   2918 catctttccggaagctcatggagcacaagcagaactcgagtttgtcggaaataattttaa 2977
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     292 CATCTTTCCGGAAGCTCATGGAGCACAAGCAGAACTCGAGTTTGTCGGAAATAATTTTAA 233

Qy   2978 tatcctgtgaaaatgaccttcatttatgccgagaatattttgccagaggcatagatgttc 3037
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    232 TATCCTGTGAAAATGACCTTCATTTATGCCGAGAATATTTTGCCAGAGGCATAGATGTTC 173

Qy   3038 acaatgcagagttcgttctgactggagtgctcactcaaacgctggactatgaatcatata 3097
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    172 ACAATGCAGAGTTCGTTCTGACTGGAGTGCTCACTCAAACGCTGGACTATGAATCATATA 113

Qy   3098 agtttaactgatggcgtctaggctgccgtgcatgtcgactcctgcggtgcggggctggct 3157
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    112 AGTTTAACTGATGGCGTCTAGGCTGCCGTGCATGTCGACTCCTGCGGTGCGGGGCTGGCT 53

Qy   3158 gtctggctggcgaggagctgctgcgcttccttcacatgctcttgttttccag 3209
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     52 GTCTGGCTGGCGAGGAGCTGCTGCGCTTCCTTCACATGCTCTTGTTTTCCAG 1


RESULT 7
AI741935/c
LOCUS       AI741935                680 bp    mRNA    linear   EST 19-DEC-1999
DEFINITION  wg29g05.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
            IMAGE:2366552 3' similar to TR:O15404 O15404 CAGF28 ;, mRNA
            sequence.
ACCESSION   AI741935
VERSION     AI741935.1  GI:5110143
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 680)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1113   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 432.
FEATURES             Location/Qualifiers
     source          1. .680
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2366552"
                     /clone_lib="Soares_NSF_F8_9W_OT_PA_P_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from five normalized
                     libraries were mixed, and ss circles were made in vitro.
                     Following HAP purification, this DNA was used as tracer in
                     a subtractive hybridization reaction. The driver was
                     PCR-amplified cDNAs from pools of 5,000 clones made from
                     the same 5 libraries. The pools consisted of the following
                     libraries and cloneIDs: Soares NbHSF pool 1:
                     309384-310919, 323208-325895 Soares Nb2HP pool 1:
                     145032-147335, 147720-148103, 148872-149255, 15002 -
                     150407, 151176-152327 Soares Nb2HF8-9W pool 1:
                     758280-760583, 772104-774407 Soares NbHPA pool 1:
                     304776-306311, 320136-322823, 326280-326663 Soares NbHOT
                     pool 1:  723720-726407, 739080-740999 Subtraction by Bento
                     Soares and M. Fatima Bonaldo."
BASE COUNT      225 a    138 c    116 g    201 t
ORIGIN


  Query Match             17.9%;  Score 639.8;  DB 9;  Length 680;
  Best Local Similarity   96.7%;  Pred. No. 4.7e-96;
  Matches  653;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy   2906 tatccaagcagccatctttccggaagctcatggagcacaagcagaactcgagtttgtcgg 2965
          |||||||||||||||||| | | ||||||||||||||||||||||||||||||||||||

Db    679 TATCCAAGCAGCCATCTTACGAGGAGCTCATGGAGCACAAGCAGAACTCGAGTTTGTCGG 620

Qy   2966 aaataatttttaatatcctgtgaaaatgaccttcatttatgccgagaatatttttgccagag 3025
          || |||| |||||||||||||| |||||||||||| |||||| |||||||| ||||||||
Db    619 AAGTAATCTTAATATCCTGTGAGAATGACCTTCATGTATGCCAAGAATATTGTGCCAGAG 560

Qy   3026 gcatagatgttcacaatgcagagttcgttctgactggagtgctcactcaaacgctggact 3085
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    559 GCATAGATGTTCACAATGCAGAGTTCGTTCTGACTGAGTGACTCACTCAAGCGCTGGACT 500

Qy   3086 atgaatcatataagtttaactgatggcgtctaggctgccgtgcatgtcgactcctgcggt 3145
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    499 ATGAATCATATAAGTTTAACTGATGGCGTCTAGGCTGCCGTGCATGTCGACTCCTGCGGT 440

Qy   3146 gcggggctggctgtctggctggcgaggagctgctgcgcttccttcacatgctcttgtttt 3205
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    439 GCGGGGCTGACTGTCTGGCTGGCGAGGAGCTGCTGCGCTTCCTTCACATGCTCTTGTTTT 380

Qy   3206 ccagctgctttcctgggggatcagactgtgaagcaggaagacagatataataaaatatact 3265
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    379 CCAGCTGCTTTCCTGGGGGATCAGACTGTGAAGCAGGAAGACAGATATAATAAATATACT 320

Qy   3266 gcatctttttaagatgtgcaattttattctgaggaaacataaattatgttttgtattata 3325
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    319 GCATCTTTTTAAGATGTGCAATTTTATTCTGAGGAAACATAAATTATGTTTTGTATTATA 260

Qy   3326 tgactttaagagcccacattaggttttatgattcatttgccaggtttttaaatgttttca 3385
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    259 TGGCTTTAAGAGCCCACATTAGGTTTTATGATTCATTTGCCAGGTTTTTAAATGTTTTCA 200

Qy   3386 caaaactgttacgggacttcaactagaaataaaatggtgtaaataaagaccttgctatct 3445
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    199 CAAAACTGTTACGGGACTTCAACTAGAAATAAAATGGTGTAAATAAAGACCTTGCTATCT 140

Qy   3446 ctaaattatggatgttaaagatttgaaatgttttgtacttgattatttttatttcttat 3505
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    139 CTAAATTATGGATGTTAAAGATTTGAAATGTTTTGTACTTTGATTATTTTTATTTCTTAT 80

Qy   3506 actctgttttctttttatattgatatcttgcccacattttaaataaatgtacttttgaact 3565
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||   |||||
Db     79 ACTCTGTTTTCTTTTATATTGATATCTTGCCCACATTTTAAATAAATGTACTAAAGAACT 20

Qy   3566 tagaaaaaaaaaaaa 3580
          || ||||||||||||
Db     19 TAAAAAAAAAAAAAA 5


RESULT 8
BG900296
LOCUS       BG900296                669 bp    mRNA    linear   EST 06-NOV-2001
DEFINITION  HOA37-1-B1 HOA (Human Osteoarthritic Cartilage) Homo sapiens cDNA,
            mRNA sequence.
ACCESSION   BG900296
VERSION     BG900296.1  GI:14310545
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 669)
  AUTHORS   Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J.,
            Sathe,G., Mui,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and
            Lark,M.W.
  TITLE     Identification and initial characterization of 5000 expressed
            sequenced tags (ESTs) each from adult human normal and
            osteoarthritic cartilage cDNA libraries
  JOURNAL   Osteoarthritis Cartilage 9 (7), 641-653 (2001)
  MEDLINE   21482651
COMMENT     Contact: Sanjay Kumar
            UW2109
            GlaxoSmithKline
            709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA
            Tel: 610-270-7245
```

```
            Fax: 610-270-5598
            Email: sanjay_kumar-1@gsk.com
            Seq primer: T7.
FEATURES              Location/Qualifiers
     source           1. .669
                      /organism-"Homo sapiens"
                      /db_xref-"taxon:9606"
                      /clone_lib-"HOA (Human Osteoarthritic Cartilage)"
                      /tissue_type="cartilage"
                      /lab_host="E.coli DH10 B"
                      /note="Vector: pSPORT I; Site_1: SalI; Site_2: NotI;
                      Directional"
BASE COUNT      200 a    179 c    159 g    130 t      1 others
ORIGIN


  Query Match           16.3%;  Score 583;  DB 10;  Length 669;
  Best Local Similarity  99.8%;  Pred. No. 1.2e-86;
  Matches  583;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   1003 ggggaattaatgtttgatgattcttcagattcatcaccggaaaaacaggagagaaatttta 1062
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     86 GGGGAATTAATGTTTGATGATTCTTCAGATTCATCACCGGAAAAACAGGAGAGAAATTTA 145

Qy   1063 aactggaccccggccgaagtcccacagttagctgcagcaaaacgcaggctgcctcaggga 1122
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    146 AACTGGACCCCGGCCGAAGTCCCACAGTTAGCTGCAGCAAAACGCAGGCTGCCTCAGGGA 205

Qy   1123 aaggagcctgggttgattaacttgtgtgccaatgtcccacccgtcccaggtaacattttg 1182
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    206 AAGGAGCCTGGGTTGATTAACTTGTGTGCCAATGTCCCACCCGTCCCAGGTAACATTTTG 265

Qy   1183 ccccctgaggtccggggtaatttaatggctgctggacaaaacctccaaagttctgaaaga 1242
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    266 CCCCCTGAGGTCCGGGGTAATTTAATGGCTGCTGGACAAAACCTCCAAAGTTCTGAAAGA 325

Qy   1243 tcagaaatgatagctacctggagtccagctgtacggacactgaggaatattactaataat 1302
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    326 TCAGAAATGATAGCTACCTGGAGTCCAGCTGTACGGACACTGAGGAATATTACTAATAAT 385

Qy   1303 gctgacattcagcagatgaaccggccatcaaatgtagcacatatcttacagactctttca 1362
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    386 GCTGACATTCAGCAGATGAACCGGCCATCAAATGTAGCACATATCTTACAGACTCTTTCA 445

Qy   1363 gcacctacgaaaaatttagaacagcaggtgaatcacagccagcagggacatacaaatgcc 1422
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    446 GCACCTACGAAAAATTTAGAACAGCAGGTGAATCACAGCCAGCAGGGACATACAAATGCC 505

Qy   1423 aatgcagtgctgtttagccaagtgaaagtgactccagagacacacatgctacagcagcag 1482
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    506 AATGCAGTGCTGTTTAGCCAAGTGAAAGTGACTCCAGAGACACACATGCTACAGCAGCAG 565

Qy   1483 cagcaggcccagcagcagcagcagcagcacccggttttacacttcagccccagcagata 1542
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    566 CAGCAGGCCCAGCAGCAGCAGCAGCAGCACCCGGTTTTACACCTTCAGCCCCAGCAGATA 625

Qy   1543 atgcagctccagcagcagcagcagcagcagatctctcagcaacc 1586
           ||||||| ||||||||||||||||||||||||||||||||||||
Db    626 ATGCAGCTNCAGCAGCAGCAGCAGCAGCAGATCTCTCAGCAACC 669


RESULT   9
AU123803
LOCUS       AU123803                625 bp    mRNA    linear   EST 23-OCT-2000
DEFINITION  AU123803 NT2RM2 Homo sapiens cDNA clone NT2RM2001075 5', mRNA
            sequence.
ACCESSION   AU123803
VERSION     AU123803.1  GI:10948519
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 625)
  AUTHORS   Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,
            Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and
            Isogai,T.
  TITLE     HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,
            Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki
            ,Y., Sugano,S., Isogai,T.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3951
            Fax: 81-438-52-3952
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES              Location/Qualifiers
     source           1. .625
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="NT2RM2001075"
                      /clone_lib="NT2RM2"
                      /cell_type="teratocarcinoma"
                      /cell_line="NT2"
                      /note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal
                      precursor cells"
BASE COUNT      198 a    154 c    141 g    126 t      6 others
ORIGIN


  Query Match           16.2%;  Score 581.2;  DB 9;  Length 625;
  Best Local Similarity  98.5%;  Pred. No. 2.3e-86;
  Matches  605;  Conservative    0;  Mismatches    7;  Indels    2;  Gaps    2;

Qy    941 aagaagggtctccttcaggtgaccagcagttttcacctaaatccaacactgaaaaatcta 1000
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2 AAGAAGC TCTCCTTCAGGTGACCAGCAGTTTTCACCTAAATCCAACACTGAAAAATCTA 61

Qy   1001 aaggggaattaatgtttgatgattcttcagattcatcaccggaaaaacaggagagaaatt 1060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     62 AAGGGGAATTAATGTTTGATGATTCTTCAGATTCATCACCGGAAAAACAGGAGAGAAATT 121

Qy   1061 taaactggaccccggccgaagtcccacagttagctgcagcaaaacgcaggctgcctcagg 1120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122 TAAACTGGACCCCGGCCGAAGTCCCACAGTTAGCTGCAGCAAAACGCAGGCTGCCTCAGG 181

Qy   1121 gaaaggagcctgggttgattaacttgtgtgccaatgtcccacccgtcccaggtaacattt 1180
          |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db    182 GAAAGGAGCCTGGGTTGATTAACTTGAGTGCCAATGTCCCACCCGTCCCAGGTAACATTT 241

Qy   1181 tgccccctgaggtccggggtaatttaatggctgctggacaaaacctccaaagttctgaaa 1240
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    242 TGCCCCCTGAGGTCCGGGGTAATTTAATGGCTGCTGGACAAAACCTGCAAAGTTCTGAAA 301

Qy   1241 gatcagaaatgata-gctacctggagtccagctgtacggacactgaggaatattactaat 1299
          |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db    302 GATCAGAAATGATATGCTACCTGGAGTCCAGCTGTACGGACACTGAGGAATATTACTAAT 361

Qy   1300 aatgctgacattcagcagatgaaccggccatcaaatgtagcacatatcttacagactctt 1359
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    362 AATGCTGACATTCAGCAGATGAACCGGCCATCAAATGTAGCACATATCTTACAGACTCTT 421

Qy   1360 tcagcacctacgaaaaatttagaacagcaggtgaatcacagccagcagggacatacaaat 1419
          |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    422 TCAGCACCTACGAAAAATTTAGAACAGCAGGTGAATCACAGNCAGCAGGGACATACAAAT 481

Qy   1420 gccaatgc-agtgctgtttagccaagtgaaagtgactccagagacacacatgctacagca 1478
          |||||||| ||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    482 GCCAATGCAAGTGCTGNTTAGCCAAGTGAAAGTGACTCCAGAGACACACATGCTACAGCA 541
```

```
Qy   1479 gcagcagcaggcccagcagcagcagcagcagcacccggttttacaccttcagccccagca 1538
          |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db    542 GCAGCAGCAGGCCCAGCAGNAGCAGCAGCAGCACCCGGTTTTACACCTTCAGCCCCAGCA 601

Qy   1539 gataatgcagctcc 1552
          ||||||||| || |
Db    602 GATAATGCANCTTC 615


RESULT 10
BE541477
LOCUS       BE541477                 803 bp    mRNA    linear   EST 09-AUG-2000
DEFINITION  601067929F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:3454183 5',
            mRNA sequence.
ACCESSION   BE541477
VERSION     BE541477.1  GI:9770122
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 803)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Incyte Genomics, Inc.
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM8438  row: j  column: 08
            High quality sequence stop: 627.
FEATURES             Location/Qualifiers
     source          1. .803
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3454183"
                     /clone_lib-"NIH_MGC_10"
                     /cell_line-"MGC36"
                     /lab_host-"DH10B"
                     /note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.5 kb. Library prepared by Life
                     Technologies."
BASE COUNT      227 a    188 c    207 g    181 t
ORIGIN


  Query Match           16.0%;  Score 573.2;  DB 10;  Length 803;
  Best Local Similarity   90.1%;  Pred. No. 4.6e-85,
  Matches  726;  Conservative    0;  Mismatches   68;  Indels   12;  Gaps   10;

Qy   2009 agccatgttcacagcatattatttctgtgactggatttgttgatagtgacagagatgacc 2068
          ||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db      1 AGCCATGTTCACAGCATATTATTTCTGTGACTGGA-TTGTTGATAGTGACAGAGATGACC 59

Qy   2069 taaaattaatggcttatttggcaggtgccaaatatacgggttatctatgccgcagcaaca 2128
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     60 TAAAATTAATGGCTTATTTGGCAGGTGCCAAATATACGGGTTATCTATGCCGCAGCAACA 119

Qy   2129 cagtcctcatctgtaaagaaccaactggcttaaagtatgaaaaagccaaagagtggagga 2188
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    120 CAGTCCTCATCTGTAAAGAACCAACTGGTTTAAAGTATGAAAAAGCCAAAGAGTGGAGGA 179

Qy   2189 taccc-tgtgtcaacgcccagtggcttggcgccattcttctgggaaactttgaggcactg 2247
          ||||| ||||||||||||| || ||||||||||||||||||||||||||||||||||||
Db    180 TACCCGTGTGTCAACGCCCA--GGGCTGGCGACATTCTTCTGGGAAACTTTGAGGCACTG 237

Qy   2248 aggcagattcagtatagtcgctacacggcattcagtctgcaggatccatttgcccctacc 2307
          ||||||| ||||||||||||||||||||||||||||||||||||||||| |||||||||
Db    238 AGGCAGATTCAGTATAGTCGCTACACGGCATTCAGTCTGCAGGATCCATTGGCCCCTACC 297

Qy   2308 cagcatttagttttaaatcttttagatgcttggagagttcccttaaaagtgtctgcagag 2367
          ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    298 CAGCATTTAGTTTTAAATCTTTTAGATGC-TGGAGAGTTCCCTTAAAAGTGTCTGCAGAG 356

Qy   2368 ttgttgatgagtataagactacctcccaaactga-aacagaatgaagtagctaatgtcca 2426
          |||||||||||||||||||||||||||||||| |||||||||||| ||||||||||||
Db    357 TTGTTGATGAGTATAAGACTACCTCCCAAACTGAGAACAGAATGAAGTAGCTAATGTCCA 416

Qy   2427 gccttcttcccaaagagccagaattgaagacgtaccacctcccactaaaaagctaactcc 2486
          |||||||||| |||||||||||||| ||||||||||||||||||||||||||||||||||
Db    417 GCCTTCTTCCAAAAGAGCCAGAATGGAAGACGTACCACCTCCCACTAAAAAGCTAACTCC 476

Qy   2487 agaattgacccctttttgtgcttttcactggattcgagcctgtccaggttcaacagtatat 2546
          ||||| |||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    477 AGAATGGACCCC-TTTGTGCTTTTCACTGGATTCGAGCCTGTCCAGGTTCAACAGTATAT 535

Qy   2547 taagaagctctacattcttggtggagaggttgcggagtctgcacagaagtgcacacac-c 2605
          |||||||||||||||||||||||||||||| |||||||||||||||||    ||||| |
Db    536 TAAGAAGCTCTACATTCTTGGTGGAGAGGTCGCGGAGTCTGCACAGAAGGTGCCACACAC 595

Qy   2606 tcattgccagcaaagtgactcgcaccgtgaagttcctgacggcgatttctgtcgtgaagc 2665
          || |||||||||||||||| |||||||||||    ||||||||| | ||||||||
Db    596 CTCATGGCAGCAAAGTGACTCGCAACGTGAAGTTCCGACGGGCGATTTC-GGCGTGAAGC 654

Qy   2666 acatagtgacgccaga--gtggctggaagaatgcttcaggtgtcagaagttcattgatga 2723
          |||||||||| |||| | |||||||||||  |    || | | ||||| | | | ||
Db    655 ACATAGTGACGGCAGAAGGGGGCTGGAAGAAGCCCCACGGGGCCCGAAGTCCTTGGGGGA 714

Qy   2724 gcagaactacattctccgagatgctgaggcagaagtacttttctctttcagcttggaaga 2783
          || |||||  |||| |    | ||  |||| |||| |  ||||| | ||    |||| |
Db    715 GCCGAACTCAATTCCCGAGAAAGCGAAGGCCGAAG-GCAATTCTCCTCCAAGGGGGAAAA 773

Qy   2784 atccttaaaacgggcacacgtttctc 2809
          | ||| |||  ||||||  ||| | |
Db    774 ACCCTGAAAGGGGGCACGGGTTCCCC 799


RESULT 11
AW009159/c
LOCUS       AW009159                 597 bp    mRNA    linear   EST 08-MAR-2000
DEFINITION  ws78b09.x1 NCI_CGAP_Co3 Homo sapiens cDNA clone IMAGE:2504057 3'
            similar to TR:O15404 O15404 CAGF28 ;contains LTR5.b2 LTR5
            repetitive element ;, mRNA sequence.
ACCESSION   AW009159
VERSION     AW009159.1  GI:5857937
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 597)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D.
            , Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arraying: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 854   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 370.
FEATURES             Location/Qualifiers
     source          1. .597
```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2504057"
/clone_lib="NCI_CGAP_Co3"
/sex="pooled"
/tissue_type="colon"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from 12 pooled bulk tumor samples and primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. "
BASE COUNT 210 a 116 c 106 g 165 t
ORIGIN

Query Match 15.6%; Score 558.4; DB 9; Length 597;
Best Local Similarity 97.1%; Pred. No. 1.4e-82;
Matches 579; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

```
Qy   2975 taatatcctgtgaaaatgaccttcatttatgccgagaatattttgccagaggcatagatg 3034
          ||| | |||||||  |   ||||| |||| ||| | ||||||||||| |||||||| |||
Db    596 TAAGACCCTGTGAGCAGACCCTTCCTTTAAGCCAAAAATATTTTGCCACAGGCATAAATG 537

Qy   3035 ttcacaatgcagagttcgttctgactggagtgctcactcaaacgctggactatgaatcat 3094
          ||||||||||||||| |||||||||||| ||||||||||||||||||||||||||||||
Db    536 TTCACAATGCAGAGATCGTTCTGACTGAAGTGCTCACTCAAACGCTGGACTATGAATCAT 477

Qy   3095 ataa-gtttaactgatggcgtctaggctgccgtgcatgtcgactcctgcggtgcggggct 3153
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    476 ATAAGGTTTAACTGATGGCGTCTAGGCTGCCGTGCATGTCGACTCCTGCGGTGCGGGGCT 417

Qy   3154 ggctgtctggctggcgaggagctgctgcgcttccttcacatgctcttgttttccagctgc 3213
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    416 GGCTGTCTGGCTGGCGAGGAGCTGCTGCGCTTCCTTCACATGCTCTTGTTTTCCAGCTGC 357

Qy   3214 tttcctgggggatcagactgtgaagcaggaagacagatataataaatatactgcatcttt 3273
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    356 TTTCCTGGGGGATCAGACTGTGAAGCAGGAAGACAGATATAATAAATATACTGCATCTTT 297

Qy   3274 ttaagatgtgcaattttattctgaggaaacataaattatgttttgtattatatgacttta 3333
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    296 TTAAGATGTGCAATTTTATTCTGAGGAAACATAAATTATGTTTTGTATTATATGACTTTA 237

Qy   3334 agagcccacattaggttttatgattcatttgccaggtttttaaatgttttcacaaaactg 3393
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    236 AGAGCCCACATTAGGTTTTATGATTCATTTGCCAGGTTTTTAAATGTTTTCACAAAACTG 177

Qy   3394 ttacgggacttcaactagaaataaaatggtgtaaataaagaccttgctatctctaaatta 3453
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    176 TTACGGGACTTCAACTAGAAATAAAATGGTGTAAATAAAGACCTTGCTATCTCTAAATTA 117

Qy   3454 tggatgttaaagatttgaaatgttttgtactttgattatttttatttcttatactctgtt 3513
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    116 TGGATGTTAAAGATTTGAAATGTTTTGTACTTTGATTATTTTTATTTCTTATACTCTGTT 57

Qy   3514 ttcttttatattgatatcttgcccacattttaaataaatgtacttttgaacttaga 3569
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db     56 TTCTTTTATATTGATATCTTGCCCACATTTTAAATAAATGTACTTTTGAACTTAAA 1
```

RESULT 12
AW004836/c
LOCUS AW004836 552 bp mRNA linear EST 10-SEP-1999
DEFINITION ws93c05.x1 NCI_CGAP_Co3 Homo sapiens cDNA clone IMAGE:2505512 3' similar to TR:O15404 O15404 CAGF28 ;contains LTR5.b2 LTR5 LTR5 repetitive element ;, mRNA sequence.
ACCESSION AW004836
VERSION AW004836.1 GI:5853614
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 552)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 390.
FEATURES Location/Qualifiers
source 1. .552
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2505512"
/clone_lib="NCI_CGAP_Co3"
/sex="pooled"
/tissue_type="colon"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from 12 pooled bulk tumor samples and primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. "
BASE COUNT 201 a 112 c 94 g 145 t
ORIGIN

Query Match 15.3%; Score 547.2; DB 9; Length 552;
Best Local Similarity 99.5%; Pred. No. 9.9e-81;
Matches 549; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy   3016 tttgccagaggcatagatgttcacaatgcagagttcgttctgactggagtgctcactcaa 3075
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    552 TTTGCCAGAGGCATAGATGTTCACAATGCAGAGTTCGTTCTGACTGGAGTGCTCACTCAA 493

Qy   3076 acgctggactatgaatcatataagtttaactgatggcgtctaggctgccgtgcatgtcga 3135
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    492 ACGCTGGACTATGAATCATATAAGTTTAACTGATGGCGTCTAGGCTGCCGTGCATGTCGA 433

Qy   3136 ctcctgcggtgcggggctggctgtctggctggcgaggagctgctgcgcttccttcacatg 3195
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    432 CTCATGCGGTGCGGGGCTGGCTGTCTGGCTGGCGAGGAGCTGCTGCGCTTCCTTCACATG 373

Qy   3196 ctcttgttttccagctgctttcctgggggatcagactgtgaagcaggaagacagatataa 3255
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    372 CTCTTGTTTTCCAGCTGCTTTCCTGGGGGATCAGACTGTGAAGCAGGAAGACAGATATAA 313

Qy   3256 taaatatactgcatctttttaagatgtgcaattttattctgaggaaacataaattatgtt 3315
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    312 TAAATATACTGCATCTTTTTAAGATGTGCAATTTTATTCTGAGGAAACATAAATTATGTT 253

Qy   3316 ttgtattatatgactttaagagcccacattaggttttatgattcatttgccaggttttta 3375
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    252 TTGTATTATATGACTTTAAGAGCCCACATTAGGTTTTATGATTCATTTGCCAGGTTTTTA 193

Qy   3376 aatgttttcacaaaactgttacgggacttcaactagaaataaaatggtgtaaataaagac 3435
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    192 AATGTTTTCACAAAACTGTTACGGGACTTCAACTAGAAATAAAATGGTGTAAATAAAGAC 133
```

```
Qy   3436 cttgctatctctaaattatggatgttaaagatttgaaatgttttgtactttgattatttt 3495
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    132 CTTGCTATCTCTAAATTATGGATGTTAAAGATTTGAAATGTTTTGTACTTTGATTATTTT 73

Qy   3496 tatttcttatactctgttttcttttatattgatatcttgcccacattttaaataaatgta 3555
          ||||||||||||||||||||||||||||||||||||||||||||||||||||  |||||
Db     72 TATTTCTTATACTCTGTTTTCTTTTATATTGATATCTTGCCCACATTTTAAATCCATGTA 13

Qy   3556 cttttgaactta 3567
          ||||||||||||
Db     12 CTTTTGAACTTA 1


RESULT 13
BF939756/c
LOCUS       BF939756                542 bp    mRNA    linear   EST 22-JAN-2001
DEFINITION  nac83a11.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:3440924 3'
            similar to TR:O15404 O15404 CAGF28 ;contains LTR5.b2 LTR5
            repetitive element ;, mRNA sequence.
ACCESSION   BF939756
VERSION     BF939756.1  GI:12356999
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 542)
  AUTHORS   NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGAP), Tumor Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
            Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Seq primer: -40UP from Gibco
            High quality sequence stop: 475.
FEATURES             Location/Qualifiers
     source          1. .542
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone-"IMAGE:3440924"
                     /clone_lib-"NCI_CGAP_Brn23"
                     /tissue_type-"glioblastoma (pooled)"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATATCTTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library is normalized, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT      198 a    109 c     90 g    145 t
ORIGIN


  Query Match           15.1%;  Score 540.4;  DB 10;  Length 542;
  Best Local Similarity   99.8%;  Pred. No. 1.3e-79;
  Matches  541;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   3027 catagatgttcacaatgcagagttcgttctgactggagtgctcactcaaacgctggacta 3086
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db    542 CATAGATGTTCACAATGCAGAGTTCGTTCTGACTGGAGTGCTCACTCAAACGCTGGACTA 483

Qy   3087 tgaatcatataagtttaactgatggcgtctaggctgccgtgcatgtcgactcctgcggtg 3146
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    482 TGAATCATATAAGTTTAACTGATGGCGTCTAGGCTGCCGTGCATGTCGACTCCTGCGGTG 423

Qy   3147 cggggctggctgtctggctggcgaggagctgctgcgcttccttcacatgctcttgttttc 3206
          ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    422 CGGGGCTGGCTGTCTGGCTGGCGAGGAGATGCTGCGCTTCCTTCACATGCTCTTGTTTTC 363

Qy   3207 cagctgctttcctgggggatcagactgtgaagcaggaagacagatataataaaatatactg 3266
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    362 CAGCTGCTTTCCTGGGGGATCAGACTGTGAAGCAGGAAGACAGATATAATAAATATACTG 303

Qy   3267 catctttttaaqatgtgcaattttattctgaggaaacataaattatgttttgtattatat 3326
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    302 CATCTTTTTAAGATGTGCAATTTTATTCTGAGGAAACATAAATTATGTTTTGTATTATAT 243

Qy   3327 gactttaagagcccacattaggttttatgattcatttgccaggtttttaaatgttttcac 3386
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    242 GACTTTAAGAGCCCACATTAGGTTTTATGATTCATTTGCCAGGTTTTTAAATGTTTTCAC 183

Qy   3387 aaaactgttacgggacttcaactagaaataaaatggtgtaaataaagaccttgctatctc 3446
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    182 AAAACTGTTACGGGACTTCAACTAGAAATAAAATGGTGTAAATAAAGACCTTGCTATCTC 123

Qy   3447 taaattatggatgttaaagatttgaaatgttttgtactttgattatttttatttcttata 3506
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122 TAAATTATGGATGTTAAAGATTTGAAATGTTTTGTACTTTGATTATTTTTATTTCTTATA 63

Qy   3507 ctctgttttcttttatattgatatcttgcccacattttaaataaatgtacttttgaacttt 3566
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     62 CTCTGTTTTCTTTTATATTGATATCTTGCCCACATTTTAAATAAATGTACTTTTGAACTT 3

Qy   3567 ag 3568
          ||
Db      2 AG 1


RESULT 14
AI807116/c
LOCUS       AI807116                598 bp    mRNA    linear   EST 18-DEC-1999
DEFINITION  we85g11.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:2347940 3' similar to TR:O15404 O15404 CAGF28 ;contains
            LTR5.b2 LTR5 repetitive element ;, mRNA sequence.
ACCESSION   AI807116
VERSION     AI807116.1  GI:5393682
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 598)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1375  Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 408.
FEATURES             Location/Qualifiers
     source          1. .598
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone-"IMAGE:2347940"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
```

Equal amounts of plasmid DNA from three normalized libraries (fetal lung NbHL19W, testis NHT, and B-cell NCI_CGAP_GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 215 a 119 c 105 g 159 t
ORIGIN

Query Match 15.1%; Score 539.6; DB 9; Length 598;
Best Local Similarity 95.0%; Pred. No. 1.8e-79;
Matches 568; Conservative 0; Mismatches 29; Indels 1; Gaps 1;

```
Qy   2971 attttaatatcctgtgaaaatgaccttcatttatgccgagaatattttgccagaggcata 3030
          ||||||| ||||||||||||||| | |   ||||| |  |||||||||||    ||||||
Db    598 ATTTTAAAATCCTGTGAAAATGCCTTCATTTTATCCGAAGAATATTTTGCAGAGGGCATA 539

Qy   3031 gatgttcacaatgcagag-ttcgttctgactggagtgctcactcaaacgctggactatga 3089
          |||||||||||||||||| |||||||||||||||||||||||||| ||||||||||||||
Db    538 GATGTTCACAATCCAGAGTTTCGTTCTGACTGGAGTGCTCACTCAACCGCTGGACTATGA 479

Qy   3090 atcatataagtttaactgatggcgtctaggctgccgtgcatgtcgactcctgcggtgcgg 3149
          ||||||||||||||||   ||||||||||||||||||||||||||    |||||||
Db    478 ATCATATAAGTTTAACTATGGGCGTCTAGGCTGCCGTGCATGTCGACTCTGCGGGTGCGG 419

Qy   3150 ggctggctgtctggctggcgaggagctgctgcgcttccttcacatgctcttgttttccag 3209
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    418 GGCTGGCTGTCTGGCTGGCGAGGAGCTGCTGCGCTTCCTTCACATGCTCTTGTTTTCCAG 359

Qy   3210 ctgctttcctgggggatcagactgtgaagcaggaagacagatataataaaatatactgcat 3269
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    358 CTGCTTTCCTGGGGGATCAGACTGTGAAGCAGGAAGACAGATATAATAAATATACTGCAT 299

Qy   3270 ctttttaagatgtgcaattttattctgaggaaacataaattatgttttgtattatatgac 3329
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    298 CTTTTTAAGATGTGCAATTTTATTCTGAGGAAACATAAATTATGTTTTGTATTATATGAC 239

Qy   3330 tttaagagcccacattaggttttatgattcatttgccaggtttttaaatgttttcacaaa 3389
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    238 TTTAAGAGCCCACATTAGGTTTTATGATTCATTTGCCAGGTTTTTAAATGTTTTCACAAA 179

Qy   3390 actgttacgggacttcaactagaaataaaatgtgtaaataaagaccttgctatctctaa 3449
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    178 ACTGTTACGGCACTTCAACTAGAAATAAAATGGTGTAAATAAAGACCTTGCTATCTCTAA 119

Qy   3450 attatggatgttaaagatttgaaatgtttctactttgattatttttatttcttatactc 3509
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    118 ATTATGGATGTTAAAGATTTGAAATGTTTTGTACTTTGATTATTTTTATTTCTTACACTC 59

Qy   3510 tgttttcttttatattgatatcttgcccacattttaaataaatgtacttttgaactta 3567
          ||||||||||||||   |||||||||| ||||||||||||||| ||||||| || ||
Db     58 TGTTTTCTTTTATATCCCTATCTTGCCCCCATTTTAAATAAATGAACTTTTGGACCTA 1
```

RESULT 15
AI393049/c
LOCUS AI393049 551 bp mRNA linear EST 30-MAR-1999
DEFINITION tg25c09.x1 NCI_CGAP_CLL1 Homo sapiens cDNA clone IMAGE:2109808 3' similar to TR:O15404 O15404 CAGF28 ;contains LTR5.b2 LTR5 repetitive element ;, mRNA sequence.
ACCESSION AI393049
VERSION AI393049.1 GI:4222596
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 551)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever, M.D., Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 738 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 359.
FEATURES Location/Qualifiers
source 1. .551
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2109808"
/clone_lib="NCI_CGAP_CLL1"
/tissue_type="B-cell, chronic lymphotic leukemia"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCATTGCTTTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized, and was constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 202 a 108 c 103 g 138 t
ORIGIN

Query Match 14.8%; Score 528.6; DB 9; Length 551;
Best Local Similarity 97.5%; Pred. No. 1.2e-77;
Matches 537; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

```
Qy   3016 tttgccagaggcatagatgttcacaatgcagagttcgttctgactggagtgctcactcaa 3075
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    551 TTTGCCAGAGGCATAGATGTTCACAATGCAGAGTTCGTTCTGACTGGAGTGCTCACTCAA 492

Qy   3076 acgctggactatgaatcatataagtttaactgatggcgtctaggctgccgtgcatgtcga 3135
          ||||||||||||  ||||||||||||||||||||||||||||||||||| ||||||||||||
Db    491 CCGCTGGACTATATATCATATAAGTTTAACTGATGGCGTCTAGGCTCCCGTGCATGTCGA 432

Qy   3136 ctcctgcggtgcggggctggctgtctggctggcgaggaggctgctgcgcttccttcacatg 3195
          |||||||||||||||||||||||||||||||| || ||||||||||||||||||||||||
Db    431 CTCCTGCGGTGCGGGGCTGGCTGTCTGGCTGGCAAGAAGCTGCTGCGCTTCCTTCACATG 372

Qy   3196 ctcttgttttccagctgctttcctgggggatcagactgtgaagcaggaagacagatataa 3255
          |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db    371 CTCTTGTTTTCCAGCTGCTTTCCTGGGGGATCAGACTGTGAAGCAGGAAGCCAGATATAA 312

Qy   3256 taaatatactgcatctttttaagatgtgcaatttttattctgaggaaacataaattatgtt 3315
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    311 TAAATATCCTGCATCTTTTTAAGATGTGCAATTTTATTCTGAGGAAACATAAATTATGTT 252

Qy   3316 ttgtattatatgactttaagagcccacattaggttttatgattcatttgccaggttttta 3375
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    251 TTGTATTATATGCCTTTAAGAGCCCACATTAGGTTTTATGATTCATTTGCCAGGTTTTTA 192

Qy   3376 aatgttttcacaaaactgttacgggacttcaactagaaataaaatggtgtaaataaagac 3435
          ||||||||| |||| |||||||||| ||||||||||||||||||||||||||||||| |
Db    191 AATGTTTTCCCAAACCTGTTACGGGCCTTCAACTAGAAATAAAATGGTGTAAATAAAGCC 132

Qy   3436 cttgctatctctaaattatggatgttaaagatttgaaatgttttgtactttgattatttt 3495
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    131 CTTGCTATCTCTAAATTATGGATGTTAAAGATTTGAAATGTTTTGTACTTTGATTATTTT 72
```

```
Qy   3496 tatttcttatactctgttttcttttatattgatatcttgcccacattttaaataaatgta 3555
          ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db     71 TATTTCTTATACTCTGTTTTCTTTTATATTGATATCTTGCCCCCATTTTAAATAAATGTA 12

Qy   3556 cttttgaactt 3566
          ||||||||||||
Db     11 CTTTTGAACTT 1


Search completed: June  4, 2002, 13:48:17
Job time: 9535 sec
```

OM nucleic - nucleic search, using sw model

Run on: June 4, 2002, 12:40:22 ; Search time 120.83 Seconds
(without alignments)
7277.729 Million cell updates/sec

Title: US-09-664-641-10
Perfect score: 3580
Sequence: 1 catagacagtcgatacgcct.....gaacttagaaaaaaaaaaaa 3580

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	110.6	3.1	397	3	US-09-253-691-3	Sequence 3, Appli
	2	103	2.9	234	1	US-08-469-802B-3	Sequence 3, Appli
	3	103	2.9	234	2	US-08-267-803B-3	Sequence 3, Appli
	4	100.2	2.8	477	4	US-09-135-994-1	Sequence 1, Appli
	5	94.8	2.6	203	4	US-09-043-303-7	Sequence 7, Appli
	6	92.6	2.6	543	6	5273901-6	Patent No. 5273901
	7	92.4	2.6	195	1	US-08-469-802B-2	Sequence 2, Appli
	8	92.4	2.6	195	2	US-08-267-803B-2	Sequence 2, Appli
	9	88.2	2.5	165	4	US-09-043-303-17	Sequence 17, Appl
	10	80.6	2.3	10348	2	US-08-457-273B-41	Sequence 41, Appl
	11	80.6	2.3	10348	3	US-08-556-419-13	Sequence 13, Appl
	12	80.6	2.3	10348	4	US-09-041-886-14	Sequence 14, Appl
	13	80.6	2.3	10366	1	US-08-246-982A-5	Sequence 5, Appli
	14	80.6	2.3	10366	1	US-08-453-265-5	Sequence 5, Appli
	15	78	2.2	154	1	US-08-469-802B-6	Sequence 6, Appli
	16	78	2.2	154	2	US-08-267-803B-6	Sequence 6, Appli
	17	77.2	2.2	533	6	5482709-5	Patent No. 5482709
	18	77	2.2	3489	2	US-08-728-323A-1	Sequence 1, Appli
c	19	77	2.2	32207	2	US-08-770-379-20	Sequence 20, Appl
c	20	77	2.2	32207	4	US-08-757-669A-20	Sequence 20, Appl
c	21	77	2.2	32207	4	US-09-230-371A-20	Sequence 20, Appl
	22	76.6	2.1	2190	4	US-09-625-188-19	Sequence 19, Appl
	23	74.6	2.1	168	1	US-08-469-802B-4	Sequence 4, Appli
	24	74.6	2.1	168	2	US-08-267-803B-4	Sequence 4, Appli
	25	74.6	2.1	171	1	US-08-469-802B-5	Sequence 5, Appli
	26	74.6	2.1	171	2	US-08-267-803B-5	Sequence 5, Appli
c	27	70.4	2.0	2214	3	US-08-864-038A-1	Sequence 1, Appli
c	28	70.4	2.0	3331	3	US-08-864-038A-2	Sequence 2, Appli
c	29	70.4	2.0	3331	3	US-08-864-038A-4	Sequence 4, Appli
	30	69.4	1.9	688	4	US-08-998-416-915	Sequence 915, App
	31	68.2	1.9	336	2	US-07-814-220-3	Sequence 3, Appli
	32	68.2	1.9	336	2	US-07-812-421-3	Sequence 3, Appli
	33	67.2	1.9	3292	1	US-07-814-964-12	Sequence 12, Appl
	34	67.2	1.9	3292	1	US-08-258-442-12	Sequence 12, Appl
	35	67.2	1.9	3292	1	US-08-328-809-7	Sequence 7, Appli
	36	67.2	1.9	3292	5	PCT-US92-11107-12	Sequence 12, Appl
	37	66.6	1.9	336	2	US-07-814-220-4	Sequence 4, Appli
c	38	66.6	1.9	336	2	US-07-814-220-4	Sequence 4, Appli
	39	66.6	1.9	336	2	US-07-812-421-4	Sequence 4, Appli
c	40	66.6	1.9	336	2	US-07-812-421-4	Sequence 4, Appli
	41	65.4	1.8	3366	1	US-08-469-802B-1	Sequence 1, Appli
	42	65.4	1.8	3366	2	US-08-267-803B-1	Sequence 1, Appli
c	43	65	1.8	336	2	US-07-814-220-3	Sequence 3, Appli
c	44	65	1.8	336	2	US-07-812-421-3	Sequence 3, Appli
	45	64	1.8	3376	1	US-08-320-559-29	Sequence 29, Appl

ALIGNMENTS

RESULT 1
US-09-253-691-3
; Sequence 3, Application US/09253691
; Patent No. 6124100
; GENERAL INFORMATION:
; APPLICANT: Dong Kyu JIN
; TITLE OF INVENTION: Diagnostic Method and Kit for Neuropsychiatric Diseases
; TITLE OF INVENTION: Using Trinucleotide Repeats Sequence
; FILE REFERENCE: 1942/36
; CURRENT APPLICATION NUMBER: US/09/253,691
; CURRENT FILING DATE: 1999-02-22
; EARLIER APPLICATION NUMBER: KR 98-6,278
; EARLIER FILING DATE: 1996-02-26
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: WordPerfect 6.1/Windows
; SEQ ID NO 3
; LENGTH: 397
; TYPE: DNA
; ORGANISM: human
US-09-253-691-3

Query Match 3.1%; Score 110.6; DB 3; Length 397;
Best Local Similarity 63.2%; Pred. No. 2.5e-18;
Matches 170; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

```
Qy   1435 tttagccaagtgaaagtgactccagagacacacatgctacagcagcagcagcaggcccag 1494
          || || ||   | ||  |   | | ||   ||   ||  ||||||||||||||   |||
Db    113 ttcagacagcagcaaaagcagcaacagcagcagcagcagcagcagcagcagcagcaq 172

Qy   1495 cagcagcagcagcagcacccggttttacaccttcagccccagcagataatgcagctccag 1554
          ||||||||||||||||| | |      || |  |||| ||||||     |||||  |||
Db    173 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 232

Qy   1555 cagcagcagcagcagcagatctctcagcaaccttaccccccagcagccgccgcatccattt 1614
          ||||||||||||||||||      |||||  |   | |  ||||||| || ||| |
Db    233 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 292

Qy   1615 tcacagcagcagcagcagcaacagcagccaccaccatcgcctcagcagcatcagcttttt 1674
            |||||||||||||||||| |||||||  |  |    ||  |||||||| ||||
Db    293 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 352

Qy   1675 ggacatgatccagcagtggagattccaga 1703
            |  || | | |||   ||| | || |
Db    353 cagcgggacctatcagcacagagttcaca 381
```

RESULT 2
US-08-469-802B-3
; Sequence 3, Application US/08469802B

; Patent No. 5741645
; GENERAL INFORMATION:
; APPLICANT: Orr, Harry T.
; APPLICANT: Ranum, Laura P.W.
; APPLICANT: Chung, Ming-yi
; APPLICANT: Zoghbi, Huda Y.
; TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia
; Patent No. 5741645
; TITLE OF INVENTION: Type 1 and Method for Diagnosis
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mueting, Raasch, Gebhardt & Schwappach, P.A.
; STREET: 119 No. 5741645th Fourth Street, Suite 203
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,802B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mueting, Ann M.
; REGISTRATION NUMBER: 33,977
; REFERENCE/DOCKET NUMBER: 110.00030101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-305-1217
; TELEFAX: 612-305-1225
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 234 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-469-802B-3

Query Match 2.9%; Score 103; DB 1; Length 234;
Best Local Similarity 66.4%; Pred. No. 1.5e-16;
Matches 148; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

Qy 1465 cacatgctacagcagcagcagcaggcccagcagcagcagcagcagcacccggttttacac 1524
Db 4 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 63

Qy 1525 cttcagccccagcagataatgcagctccagcagcagcagcagcagcagatctctcagcaa 1584
Db 64 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 123

Qy 1585 ccttacccccagcagccgccgcatccatttcacagcagcagcagcagcaacagcagcca 1644
Db 124 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 183

Qy 1645 ccaccatcgcctcagcagcatcagctttttggacatgatccag 1687
Db 184 CAGCAGCAGCAGCAGCAGCAGCACCTCAGCAGGGCTCCGG 226

RESULT 3
US-08-267-803B-3
; Sequence 3, Application US/08267803B
; Patent No. 5834183
; GENERAL INFORMATION:
; APPLICANT: Orr, Harry T.
; APPLICANT: Ranum, Laura P.W.
; APPLICANT: Chung, Ming-yi
; APPLICANT: Zoghbi, Huda Y.
; TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia
; Patent No. 5834183
; TITLE OF INVENTION: Type 1 and Method for Diagnosis
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mueting, Raasch, Gebhardt & Schwappach, P.A.
; STREET: P.O. Box 581415
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55458-1415
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/267,803B
; FILING DATE: 28-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McCormack, Myra H.
; REGISTRATION NUMBER: 36,602
; REFERENCE/DOCKET NUMBER: 110.00030120
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-305-1217
; TELEFAX: 612-305-1228
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 234 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-267-803B-3

Query Match 2.9%; Score 103; DB 2; Length 234;
Best Local Similarity 66.4%; Pred. No. 1.5e-16;
Matches 148; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

Qy 1465 cacatgctacagcagcagcagcaggcccagcagcagcagcagcagcacccggttttacac 1524
Db 4 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 63

Qy 1525 cttcagccccagcagataatgcagctccagcagcagcagcagcagcagatctctcagcaa 1584
Db 64 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 123

Qy 1585 ccttacccccagcagccgccgcatccatttcacagcagcagcagcagcaacagcagcca 1644
Db 124 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 183

Qy 1645 ccaccatcgcctcagcagcatcagctttttggacatgatccag 1687
Db 184 CAGCAGCAGCAGCAGCAGCAGCACCTCAGCAGGGCTCCGG 226

RESULT 4
US-09-135-994-1
; Sequence 1, Application US/09135994A
; Patent No. 6280938
; GENERAL INFORMATION:
; APPLICANT: Ranum et al.
; TITLE OF INVENTION: SCA7 GENE AND METHODS OF USE
; FILE REFERENCE: University of Minnesota
; CURRENT APPLICATION NUMBER: US/09/135,994A
; CURRENT FILING DATE: 1998-08-18
; EARLIER APPLICATION NUMBER: 60/056,170
; EARLIER FILING DATE: 1997-08-19
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1

; LENGTH: 477
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-135-994-1

Query Match 2.8%; Score 100.2; DB 4; Length 477;
Best Local Similarity 69.9%; Pred. No. 1.1e-15;
Matches 135; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

Qy 1477 cagcagcagcaggcccagcagcagcagcagcacccggttttacaccttcagccccag 1536
Db 124 cggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 183

Qy 1537 cagataatgcagctccagcagcagcagcagcagatctctcagcaaccttacccccag 1596
Db 184 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 243

Qy 1597 cagccgccgcatccatttcacagcagcagcagcagcaacagcagccaccaccatcgcct 1656
Db 244 cagcagcagcagcagcagcagcagcagcagcagcagcagccgccgcctccgcagccc 303

Qy 1657 cagcagcatcagc 1669
Db 304 cagcggcagcagc 316

RESULT 5
US-09-043-303-7
; Sequence 7, Application US/09043303
; Patent No. 6251589
; GENERAL INFORMATION:
; APPLICANT: TSUJI, Shoji
; APPLICANT: SANPEI, Kazujiro
; TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Alaxia Type 2 and
; TITLE OF INVENTION: Primers Therefor
; FILE REFERENCE: 0760-0241P
; CURRENT APPLICATION NUMBER: US/09/043,303
; CURRENT FILING DATE: 1998-05-18
; EARLIER APPLICATION NUMBER: PCT/JP96/01999
; EARLIER FILING DATE: 1996-07-18
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 203
; TYPE: DNA
; ORGANISM: p-2093 plasmid
US-09-043-303-7

Query Match 2.6%; Score 94.8; DB 4; Length 203;
Best Local Similarity 70.8%; Pred. No. 1.5e-14;
Matches 126; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Qy 1465 cacatgctacagcagcagcaggcccagcagcagcagcagcacccggttttacac 1524
Db 4 caccagcaacagcaacagcagcagcagcagcagcagcagcagcagcagcagcagcag 63

Qy 1525 cttcagccccagcagataatgcagctccagcagcagcagcagcagatctctcagcaa 1584
Db 64 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 123

Qy 1585 ccttaccccagcagccgccgcatccattttcacagcagcagcagcagcaacagcagc 1642
Db 124 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 181

RESULT 6
5273901-6
;Patent No. 5273901
; APPLICANT: JACOBSON, JAMES W.;STRAUSBERG, ROBERT L.;WILSON,
;SUSAN D.;POPE, SHARON H.;STRAUSBERG, SUSAN L.;RUFF,MICHAEL D.;
;AUGUSTINE, PATRICIA C.;DANFORTH, HARRY D.
; TITLE OF INVENTION: GENETICALLY ENGINEERED COCCIDIOSIS
;SPOROZOITE 21.5 KB ANTIGEN, AC-6B
; NUMBER OF SEQUENCES: 11
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/581,693
; FILING DATE: 12-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 215,162
; FILING DATE: 05-JUL-1988
; APPLICATION NUMBER: 746,520
; FILING DATE: 19-JUN-1985
; APPLICATION NUMBER: 627,811
; FILING DATE: 05-JUL-1984
;SEQ ID NO:6:
; LENGTH:543
5273901-6

Query Match 2.6%; Score 92.6; DB 6; Length 543;
Best Local Similarity 56.9%; Pred. No. 9.1e-14;
Matches 170; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

Qy 1369 acgaaaaatttagaacagcaggtgaatcacagccagcagggacatacaaatgccaatgca 1428
Db 5 aagaagacaaaagagaagaagaagacaaaagagaagaagaagaagaagaaagagaagaga 64

Qy 1429 gtgctgtttagccaagtgaaagtgactccagagacacacatgctacagcagcagcagcag 1488
Db 65 gaggagaagaagaagaaaaagaagaagaaagagcggcagcacctgcagcagcaacagcag 124

Qy 1489 gcccagcagcagcagcagcagcacccggttttacaccttcagccccagcagataatgcag 1548
Db 125 cagcacctgcagcagcaacagcagcagcaacagcagctgcaacagcagcagctacagcaa 184

Qy 1549 ctccagcagcagcagcagcagatctctcagcaaccttacccccagcagccgccgcat 1608
Db 185 caccagctgc.gctgcagcaccagctgcagcagcaccagctgcagcagcagcagcagcaa 244

Qy 1609 ccattttcacagcagcagcagcagcaacagcagccaccaccatcgcctcagcagcatca 1667
Db 245 caggagctgcagcagcaccagctgcagcgacagccgcagcaggagcagcaggagcagca 303

RESULT 7
US-08-469-802B-2
; Sequence 2, Application US/08469802B
; Patent No. 5741645
; GENERAL INFORMATION:
; APPLICANT: Orr, Harry T.
; APPLICANT: Ranum, Laura P.W.
; APPLICANT: Chung, Ming-yi
; APPLICANT: Zoghbi, Huda Y.
; TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia
; Patent No. 5741645
; TITLE OF INVENTION: Type 1 and Method for Diagnosis
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mueting, Raasch, Gebhardt & Schwappach, P.A.
; STREET: 119 No. 5741645th Fourth Street, Suite 203
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,802B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

```
;       NAME:  Mueting, Ann M.
;       REGISTRATION NUMBER:  33,977
;       REFERENCE/DOCKET NUMBER:  110.00030101
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  612-305-1217
;       TELEFAX:  612-305-1225
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  195 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA
US-08-469-802B-2


  Query Match             2.6%;  Score 92.4;  DB 1;  Length 195;
  Best Local Similarity   67.9%;  Pred. No. 6e-14;
  Matches  129;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy   1474 cagcagcagcagcaggcccagcagcagcagcagcagcacccggttttacaccttcagccc 1533
          ||||||||||||||||   ||||||||||||||||||||| | |      || |  ||||
Db      1 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 60

Qy   1534 cagcagataatgcagctccagcagcagcagcagcagcagatctctcagcaaccttacccc 1593
          ||||||     |||||  ||||||||||||||||||||      |||||  |  | |
Db     61 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 120

Qy   1594 cagcagccgccgcatccattttcacagcagcagcagcagcaacagcagccaccaccatcg 1653
          ||||||| || ||| |        ||||||||||||||||| |||||||  |       |
Db    121 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCACCTCAGCAGG 180

Qy   1654 cctcagcagc 1663
          ||| |  ||
Db    181 GCTCCGGGGC 190


RESULT  8
US-08-267-803B-2
; Sequence 2, Application US/08267803B
; Patent No. 5834183
;  GENERAL INFORMATION:
;    APPLICANT:  Orr, Harry T.
;    APPLICANT:  Ranum, Laura P.W.
;    APPLICANT:  Chung, Ming-yi
;    APPLICANT:  Zoghbi, Huda Y.
;    TITLE OF INVENTION:  Gene Sequence for Spinocerebellar Ataxia
; Patent No. 5834183
;    TITLE OF INVENTION:  Type 1 and Method for Diagnosis
;    NUMBER OF SEQUENCES:  85
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Mueting, Raasch, Gebhardt & Schwappach, P.A.
;      STREET:  P.O. Box 581415
;      CITY:  Minneapolis
;      STATE:  MN
;      COUNTRY:  USA
;      ZIP:  55458-1415
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/267,803B
;      FILING DATE:  28-JUN-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McCormack, Myra H.
;      REGISTRATION NUMBER:  36,602
;      REFERENCE/DOCKET NUMBER:  110.00030120
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  612-305-1217
;      TELEFAX:  612-305-1228
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  195 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-08-267-803B-2


  Query Match             2.6%;  Score 92.4;  DB 2;  Length 195;
  Best Local Similarity   67.9%;  Pred. No. 6e-14;
  Matches  129;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy   1474 cagcagcagcagcaggcccagcagcagcagcagcagcacccggttttacaccttcagccc 1533
          ||||||||||||||||   ||||||||||||||||||||| | |      || |  ||||
Db      1 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 60

Qy   1534 cagcagataatgcagctccagcagcagcagcagcagcagatctctcagcaaccttacccc 1593
          ||||||     |||||  ||||||||||||||||||||      |||||  |  | |
Db     61 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 120

Qy   1594 cagcagccgccgcatccattttcacagcagcagcagcagcaacagcagccaccaccatcg 1653
          ||||||| || ||| |        ||||||||||||||||| |||||||  |       |
Db    121 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCACCTCAGCAGG 180

Qy   1654 cctcagcagc 1663
          ||| |  ||
Db    181 GCTCCGGGGC 190


RESULT  9
US-09-043-303-17
; Sequence 17, Application US/09043303
; Patent No. 6251589
; GENERAL INFORMATION:
;  APPLICANT: TSUJI, Shoji
;  APPLICANT: SANPEI, Kazujiro
;  TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Alaxia Type 2 and
;  TITLE OF INVENTION: Primers Therefor
;  FILE REFERENCE: 0760-0241P
;  CURRENT APPLICATION NUMBER: US/09/043,303
;  CURRENT FILING DATE: 1998-05-18
;  EARLIER APPLICATION NUMBER: PCT/JP96/01999
;  EARLIER FILING DATE: 1996-07-18
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
;   LENGTH: 165
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:Probe
US-09-043-303-17


  Query Match             2.5%;  Score 88.2;  DB 4;  Length 165;
  Best Local Similarity   70.9%;  Pred. No. 6.1e-13;
  Matches  117;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy   1474 cagcagcagcagcaggcccagcagcagcagcagcagcacccggttttacaccttcagccc 1533
          ||||||||||||||||   ||||||||||||||||||||| | |      || |  ||||
Db      1 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 60

Qy   1534 cagcagataatgcagctccagcagcagcagcagcagcagatctctcagcaaccttacccc 1593
          ||||||     |||||  ||||||||||||||||||||      |||||  |  | |
Db     61 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 120

Qy   1594 cagcagccgccgcatccattttcacagcagcagcagcagcaacag 1638
          ||||||| || ||| |        ||||||||||||||||| |||
Db    121 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 165
```

```
RESULT 10
US-08-457-273B-41
; Sequence 41, Application US/08457273B
; Patent No. 5849995
;  GENERAL INFORMATION:
;    APPLICANT:  Hayden, Michael
;    APPLICANT:  Lin, Biaoyang
;    APPLICANT:  Nasir, Jamal
;    TITLE OF INVENTION:  Mouse Model for Huntington's Disease and
;    TITLE OF INVENTION:  Related DNA Sequences
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Virginia Bennett
;      STREET:  PO Box 37428
;      CITY:  Raleigh
;      STATE:  No. 5849995th Carolina
;      COUNTRY:  US
;      ZIP:  27627
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,273B
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bennett, Virginia C.
;      REGISTRATION NUMBER:  37,092
;      REFERENCE/DOCKET NUMBER:  3477-85A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-854-1400
;      TELEFAX:  919-854-1401
;  INFORMATION FOR SEQ ID NO:  41:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10348 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-457-273B-41


  Query Match             2.3%;  Score 80.6;  DB 2;  Length 10348;
  Best Local Similarity   62.8%;  Pred. No. 4.1e-10;
  Matches  125;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy   1459 gagacacacatgctacagcagcagcagcaggcccagcagcagcagcagcagcacccggtt 1518
          ||| | | || |      ||||||||||||   ||||||||||||||||||| | |
Db    349 GAGTCCCTCAAGTCCTTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 408

Qy   1519 ttacaccttcagccccagcagataatgcagctccagcagcagcagcagcagcagatctct 1578
             || |  ||||  ||||||  |  || ||  | || || || || || || |    ||
Db    409 CAGCAGCAGCAGCAGCAGCAACAGCCGCCACCGCCGCCGCCGCCGCCGCCTCCT 468

Qy   1579 cagcaaccttacccccagcagccgccgcatccattttcacagcagcagcagcagcaacag 1638
          ||||  ||| | || | || || | | || ||  |    |  |||| ||||| ||  | |
Db    469 CAGCTTCCTCAGCCGCCGCCGCAGGCACAGCCGCTGCTGCCTCAGCCGCAGCCGCCCCCG 528

Qy   1639 cagccaccaccatcgcctc 1657
          | ||| || ||  |||| |
Db    529 CCGCCGCCCCCGCCGCCAC 547


RESULT 11
US-08-556-419-13
; Sequence 13, Application US/08556419C
; Patent No. 6093549
; GENERAL INFORMATION:
;  APPLICANT: Ross, Christopher
;  APPLICANT: Li, Xiao-Jiang
;  APPLICANT: Li, Shi-Hua
;  APPLICANT: Sharp, Alan
;  APPLICANT: Lanahan, Anthony
;  APPLICANT: Worley, Paul
;  APPLICANT: Snyder, Solomon
;  TITLE OF INVENTION: Huntingtin-associated protein
;  FILE REFERENCE: 01107.52271
;  CURRENT APPLICATION NUMBER: US/08/556,419C
;  CURRENT FILING DATE: 1995-11-09
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 13
;   LENGTH: 10348
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-08-556-419-13


  Query Match             2.3%;  Score 80.6;  DB 3;  Length 10348;
  Best Local Similarity   62.8%;  Pred. No. 4.1e-10;
  Matches  125;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy   1459 gagacacacatgctacagcagcagcagcaggcccagcagcagcagcagcagcacccggtt 1518
          ||| | | || |      ||||||||||||   ||||||||||||||||||| | |
Db    349 gagtccctcaagtccttccagcagcagcagcagcagcagcagcagcagcagcagcag 408

Qy   1519 ttacaccttcagccccagcagataatgcagctccagcagcagcagcagcagcagatctct 1578
             || |  ||||  ||||||  |  || ||  | || || || || || || |    ||
Db    409 cagcagcagcagcagcagcagcaacagccgccaccgccgccgccgccgccgcctcct 468

Qy   1579 cagcaaccttacccccagcagccgccgcatccattttcacagcagcagcagcagcaacag 1638
          ||||  ||| | || | || || | | || ||  |    |  |||| ||||| ||  | |
Db    469 cagcttcctcagccgccgccgcaggcacagccgctgctgcctcagccgcagccgccccccg 528

Qy   1639 cagccaccaccatcgcctc 1657
          | ||| || ||  |||| |
Db    529 ccgccgcccccgccgccac 547


RESULT 12
US-09-041-886-14
; Sequence 14, Application US/09041886
; Patent No. 6235872
;  GENERAL INFORMATION:
;    APPLICANT:  Bredesen, Dale E.
;    APPLICANT:  Rabizadeh, Sharroz
;    TITLE OF INVENTION:  Proapoptotic Peptides, Dependence
;    TITLE OF INVENTION:  Polypeptides and Methods of Use
;    NUMBER OF SEQUENCES:  72
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell & Flores LLP
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/041,886
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LJ 2626
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
```

```
;      LENGTH:  10348 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  316..9748
US-09-041-886-14


  Query Match             2.3%;  Score 80.6;  DB 4;  Length 10348;
  Best Local Similarity   62.8%;  Pred. No. 4.1e-10;
  Matches  125;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy   1459 gagacacacatgctacagcagcagcagcaggcccagcagcagcagcagcagcacccggtt 1518
          ||| | | || |      ||||||||||||   ||||||||||||||||||||| | |
Db    349 GAGTCCCTCAAGTCCTTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 408

Qy   1519 ttacaccttcagccccagcagataatgcagctccagcagcagcagcagcagcagatctct 1578
            || |  ||||  ||||||  |  || ||  | || || || || || || |    ||
Db    409 CAGCAGCAGCAGCAGCAGCAACAGCCGCCACCGCCGCCGCCGCCGCCGCCGCCTCCT 468

Qy   1579 cagcaaccttacccccagcagccgccgcatccattttcacagcagcagcagcagcaacag 1638
          ||||  ||| | || | || || | | || ||  |    |  |||| ||||| ||  | |
Db    469 CAGCTTCCTCAGCCGCCGCCGCAGGCACAGCCGCTGCTGCCTCAGCCGCAGCCGCCCCCG 528

Qy   1639 cagccaccaccatcgcctc 1657
          | ||| || ||  |||| |
Db    529 CCGCCGCCCCCGCCGCCAC 547


RESULT  13
US-08-246-982A-5
; Sequence 5, Application US/08246982A
; Patent No. 5686288
;  GENERAL INFORMATION:
;    APPLICANT:  MacDonald, Marcy E.
;    APPLICANT:  Ambrose, Christine M.
;    APPLICANT:  Duyao, Mabel P.
;    APPLICANT:  Gusella, James F.
;    TITLE OF INVENTION:  Huntingtin DNA, Protein And Uses Thereof
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox
;      STREET:  1100 New York Avenue
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/246,982A
;      FILING DATE:  May 20, 1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Goldstein, Jorge, A.
;      REGISTRATION NUMBER:  29,021
;      REFERENCE/DOCKET NUMBER:  0609.3880002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10366 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  316..9748
US-08-246-982A-5


  Query Match             2.3%;  Score 80.6;  DB 1;  Length 10366;
  Best Local Similarity   62.8%;  Pred. No. 4.1e-10;
  Matches  125;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy   1459 gagacacacatgctacagcagcagcagcaggcccagcagcagcagcagcagcacccggtt 1518
          ||| | | || |      ||||||||||||   ||||||||||||||||||||| | |
Db    349 GAGTCCCTCAAGTCCTTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 408

Qy   1519 ttacaccttcagccccagcagataatgcagctccagcagcagcagcagcagcagatctct 1578
            || |  ||||  ||||||  |  || ||  | || || || || || || |    ||
Db    409 CAGCAGCAGCAGCAGCAGCAACAGCCGCCACCGCCGCCGCCGCCGCCGCCGCCTCCT 468

Qy   1579 cagcaaccttacccccagcagccgccgcatccattttcacagcagcagcagcagcaacag 1638
          ||||  ||| | || | || || | | || ||  |    |  |||| ||||| ||  | |
Db    469 CAGCTTCCTCAGCCGCCGCCGCAGGCACAGCCGCTGCTGCCTCAGCCGCAGCCGCCCCCG 528

Qy   1639 cagccaccaccatcgcctc 1657
          | ||| || ||  |||| |
Db    529 CCGCCGCCCCCGCCGCCAC 547


RESULT  14
US-08-453-265-5
; Sequence 5, Application US/08453265
; Patent No. 5693757
;  GENERAL INFORMATION:
;    APPLICANT:  MacDonald, Marcy E.
;    APPLICANT:  Ambrose, Christine M.
;    APPLICANT:  Duyao, Mabel P.
;    APPLICANT:  Gusella, James F.
;    TITLE OF INVENTION:  Huntingtin DNA, Protein And Uses Thereof
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox
;      STREET:  1100 New York Avenue
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/453,265
;      FILING DATE:  30-MAY-1995
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ludwig, Steven R.
;      REGISTRATION NUMBER:  36,203
;      REFERENCE/DOCKET NUMBER:  0609.3880003
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10366 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  316..9748
US-08-453-265-5


  Query Match             2.3%;  Score 80.6;  DB 1;  Length 10366;
```

```
  Best Local Similarity   62.8%;  Pred. No. 4.1e-10;
  Matches  125;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy   1459 gagacacacatgctacagcagcagcagcaggcccagcagcagcagcagcagcacccggtt 1518
          ||| | | || |      ||||||||||||   ||||||||||||||||||| | |
Db    349 GAGTCCCTCAAGTCCTTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 408

Qy   1519 ttacaccttcagccccagcagataatgcagctccagcagcagcagcagcagcagatctct 1578
             || |  ||||  ||||||  |  || ||  | || || || || || || | ||
Db    409 CAGCAGCAGCAGCAGCAGCAACAGCCGCCACCGCCGCCGCCGCCGCCGCCTCCT 468

Qy   1579 cagcaaccttacccccagcagccgccgcatccattttcacagcagcagcagcagcaacag 1638
          ||||  ||| | || | || || | | || ||  |    |  |||| ||||| ||  | |
Db    469 CAGCTTCCTCAGCCGCCGCCGCAGGCACAGCCGCTGCTGCCTCAGCCGCAGCCGCCCCCG 528

Qy   1639 cagccaccaccatcgcctc 1657
          | ||| || ||  |||| |
Db    529 CCGCCGCCCCCGCCGCCAC 547


RESULT  15
US-08-469-802B-6
; Sequence 6, Application US/08469802B
; Patent No. 5741645
;  GENERAL INFORMATION:
;    APPLICANT:  Orr, Harry T.
;    APPLICANT:  Ranum, Laura P.W.
;    APPLICANT:  Chung, Ming-yi
;    APPLICANT:  Zoghbi, Huda Y.
;    TITLE OF INVENTION:  Gene Sequence for Spinocerebellar Ataxia
; Patent No. 5741645
;    TITLE OF INVENTION:  Type 1 and Method for Diagnosis
;    NUMBER OF SEQUENCES:  47
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Mueting, Raasch, Gebhardt & Schwappach, P.A.
;      STREET:  119 No. 5741645th Fourth Street, Suite 203
;      CITY:  Minneapolis
;      STATE:  MN
;      COUNTRY:  USA
;      ZIP:  55401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/469,802B
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mueting, Ann M.
;      REGISTRATION NUMBER:  33,977
;      REFERENCE/DOCKET NUMBER:  110.00030101
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  612-305-1217
;      TELEFAX:  612-305-1225
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  154 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-08-469-802B-6


  Query Match             2.2%;  Score 78;  DB 1;  Length 154;
  Best Local Similarity   70.0%;  Pred. No. 2.1e-10;
  Matches  105;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy   1493 agcagcagcagcagcagcacccggttttacaccttcagccccagcagataatgcagctcc 1552
          |||||||||||||||||| | |      || |  ||||  ||||||     |||||  |
Db      3 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 62

Qy   1553 agcagcagcagcagcagcagatctctcagcaaccttacccccagcagccgccgcatccat 1612
          ||||||||||||||||||||      ||||| |   | |  ||||||| || ||| |
Db     63 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 122

Qy   1613 tttcacagcagcagcagcagcaacagcagc 1642
               ||||||||||||||||| |||||||
Db    123 AGCAGCAGCAGCAGCAGCAGCAGCAGC 152


Search completed: June  4, 2002, 15:36:56
Job time: 10594 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 4, 2002, 12:45:52 ; Search time 529 Seconds
(without alignments)
11619.194 Million cell updates/sec

Title: US-09-664-641-10
Perfect score: 3580
Sequence: 1 catagacagtcgatacgcct.....gaacttagaaaaaaaaaaaa 3580

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	3519	98.3	3519	22	AAS44722	Human full-length
	2	331.2	9.3	335	21	AAC27011	Human secreted pro
	3	204.6	5.7	690	22	AAH70661	Human cervical can
	4	186.8	5.2	504	22	AAH71865	Human cervical can
	5	186.8	5.2	579	22	AAH68998	Human cervical can
	6	162	4.5	467	22	AAI85135	Human polynucleoti
	7	135.6	3.8	150	19	AAX11992	Human biallelic po
c	8	125	3.5	344	22	AAS44894	Human contig polyn
	9	110.6	3.1	397	20	AAX89891	Spinocerebellar at
	10	105.6	2.9	5120	22	AAC84677	DNA sequence of hu
c	11	105.2	2.9	1037	21	AAA59242	Exons E, C and A o
c	12	105.2	2.9	1159	21	AAA59240	An EcoRI fragment
c	13	105.2	2.9	1472	21	AAA59241	Exons D, C, B and
	14	103	2.9	234	16	AAQ84832	Spinocerebellar at
	15	100.2	2.8	477	21	AAZ44307	Human SCA7 genomic
	16	94.8	2.6	203	19	AAV30271	Glutamine rich reg
	17	94.8	2.6	203	19	AAV17226	SCA2 gene CAG repe
	18	94.2	2.6	486	22	AAF75507	Polyglutamine trac
	19	93.2	2.6	6794	21	AAZ45597	cDNA sequence of h
	20	92.6	2.6	543	13	AAQ23092	Antigen tc-7a gene
	21	92.4	2.6	195	16	AAQ84831	Spinocerebellar at
c	22	89.6	2.5	567	21	AAA29550	HIV codon altered
c	23	89.4	2.5	1400	23	AAS83195	DNA encoding novel
	24	88.2	2.5	165	19	AAV30274	Glutamine rich reg
	25	86.6	2.4	6273	23	ABL15731	Drosophila melanog
c	26	86.6	2.4	10855	23	ABL15730	Drosophila melanog
c	27	85	2.4	4335	23	AAS79854	DNA encoding novel
	28	85	2.4	9732	23	AAS79690	DNA encoding novel
	29	84.8	2.4	3263	23	ABL10403	Drosophila melanog
	30	84.8	2.4	16962	23	ABL10402	Drosophila melanog
	31	83.4	2.3	4409	23	ABL17895	Drosophila melanog
	32	83.4	2.3	5021	23	ABL11783	Drosophila melanog
	33	83.4	2.3	7394	23	ABL17894	Drosophila melanog
	34	83.4	2.3	7614	23	ABL11782	Drosophila melanog
	35	83.2	2.3	10910	23	ABL03131	Drosophila melanog
	36	83	2.3	2215	24	ABI99688	Mouse ischaemic co
	37	83	2.3	6558	21	AAZ45602	cDNA sequence of a
	38	82.8	2.3	2672	23	ABL08857	Drosophila melanog
c	39	82.8	2.3	4810	23	ABL08856	Drosophila melanog
	40	82.4	2.3	14704	13	AAQ20685	PKS 741 insert con
	41	82.2	2.3	10348	22	ABA08714	Human Huntington's
	42	82	2.3	629	21	AAF08420	Fusarium venenatum
	43	81.4	2.3	1713	23	ABL18459	Drosophila melanog
c	44	81.4	2.3	5157	23	ABL18458	Drosophila melanog
c	45	81.2	2.3	3791	23	ABL04336	Drosophila melanog

ALIGNMENTS

RESULT 1
AAS44722
ID AAS44722 standard; DNA; 3519 BP.
XX
AC AAS44722;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human full-length polynucleotide sequence #147.
XX
KW Mammal; human; rhesus monkey; baker's yeast; fission yeast; Norway rat;
KW mouse; Chinese hamster; African clawed frog; fruit fly; dog; leukaemia;
KW cancer; lymphoma; neuroblastoma; autoimmune disorder; cell proliferation;
KW nervous system disorder; inflammatory disorder; cell differentiation; ds;
KW angiogenesis; stem cell growth factor; activin; inhibin; cartilage; burn;
KW genetic disorder; bone regeneration; tendon; ligament; tissue repair;
KW cytostatic; antirheumatic; antiarthritic; vulnerary; antiinflammatory;
KW antibacterial; immunosuppressive; vasotropic; antiparkinsonian;
KW neuroprotective; osteopathic; antidiabetic; antiasthmatic; antiallergic;
KW immunostimulant; analgesic; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200164834-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US04926.
XX
PR 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
PR 17-JUN-2000; 2000US-0597707.

PR 14-JUL-2000; 2000US-0616807.
PR 19-SEP-2000; 2000US-0664641.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Wang J, Ma Y, Wang D, Chen R, Xu C;
PI Drmanac R;
XX
DR WPI; 2001-589862/66.
DR P-PSDB; AAU27822.
XX
PT Novel polypeptides and nucleic acids obtained from cDNA libraries
PT prepared from various human tissues, for diagnosis, treatment of
PT cancer, neurological, inflammatory disorders and for use in arrays for
PT detection -
XX
PS Claim 1; SEQ ID No 147; 153pp; English.
XX
CC Sequences AAS44576-AAS44919 represent full-length polynucleotides and
CC contig polynucleotides encoding polypeptides of the invention. The DNA
CC and protein sequences are useful for the treatment, diagnosis and
CC prevention of various types of disorder in a mammalian subject such as a
CC human, dog, monkey, mouse, hamster or rat. The disorders include cancers
CC such as leukaemia, lymphoma and neuroblastoma, autoimmune disorders such
CC as multiple sclerosis, connective tissue disease, rheumatoid arthritis,
CC diabetes mellitus, allergic rhinitis, asthma and eczema, nervous system
CC disorders such as Parkinson's disease, Alzheimer's disease, Huntington's
CC chorea, amyotrophic lateral sclerosis, spinal muscular atrophy and
CC Wernicke disease, inflammatory disorders such as nephritis, Crohn's
CC disease, ischaemia-reperfusion injury, shock, sepsis and inflammatory
CC bowel disease. The sequences exhibit activity relating to angiogenesis,
CC cell proliferation, cell differentiation, stem cell growth factor,
CC activin or inhibin. Therefore, they can be used to manipulate stem cells
CC in culture to give rise to neuroepithelial cells that can be used to
CC augment or replace cells damaged by illness, accidental damage or genetic
CC disorders. The sequences may also be used for regeneration of bone,
CC cartilage, tendons and ligaments and in tissue repair and burn healing.
CC Note: Some sequences for this patent did not form part of the printed
CC specification, but were obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3519 BP; 996 A; 783 C; 839 G; 901 T; 0 other;

Query Match 98.3%; Score 3519; DB 22; Length 3519;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3519; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      62 agtgcaagacgggcgggctgctgagggcgccccgcgggtctggtctgggccggggaactc 121
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 agtgcaagacgggcgggctgctgagggcgccccgcgggtctggtctgggccggggaactc 60

Qy     122 cgggacggcgccgcagcagggctacggatctctcggcccctgcccttcaaatcacctcta 181
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 cgggacggcgccgcagcagggctacggatctctcggcccctgcccttcaaatcacctcta 120

Qy     182 ggcacaggttattcagcttctcaaggctggaaaagcgaaggaagtttcctacaatgcact 241
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ggcacaggttattcagcttctcaaggctggaaaagcgaaggaagtttcctacaatgcact 180

Qy     242 agcctcacacataatctcagaggatggggacaatccagaggtgggagaagctcgggaagt 301
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 agcctcacacataatctcagaggatggggacaatccagaggtgggagaagctcgggaagt 240

Qy     302 ctttgacttacctgttgtaaagccttcttgggtgattctgtccgttcagtgtggaactct 361
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 ctttgacttacctgttgtaaagccttcttgggtgattctgtccgttcagtgtggaactct 300

Qy     362 tctgccagtaaatggtttttctccagaatcatgtcagatttttttggaatcactgcctg 421
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 tctgccagtaaatggtttttctccagaatcatgtcagatttttttggaatcactgcctg 360

Qy     422 cctttctcagggtgttgatacaagctggagctctttgttggagtcttccagagctctccc 481
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 cctttctcagggtgttgatacaagctggagctctttgttggagtcttccagagctctccc 420

Qy     482 aaggagaggtagggaagggagcttgtccagcagaagttgggaagcacagagatcatctgc 541
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 aaggagaggtagggaagggagcttgtccagcagaagttgggaagcacagag tcatctgc 480

Qy     542 cttcttctgacccggtattgatgcaggctgaggcctctgttgtaatgtgctgggtgtcat 601
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 cttcttctgacccggtattgatgcaggctgaggcctctgttgtaatgtgctgggtgtcat 540

Qy     602 ctgaagacagaagtgccctgtgggctttggttacgttctatgggggagattgccagctaa 661
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 ctgaagacagaagtgccctgtgggctttggttacgttctatgggggagattgccagctaa 600

Qy     662 ccctcaataagaaatgcacgcatttgattgttccagagccaaaggggggggaaatacgaat 721
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 ccctcaataagaaatgcacgcatttgattgttccagagccaaaggggggggaaatacgaat 660

Qy     722 gtgctttaaagcgagcaagtattaaaattgtgactcctgactgggttctggattgcgtat 781
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 gtgctttaaagcgagcaagtattaaaattgtgactcctgactgggttctggattgcgtat 720

Qy     782 cagagaaaaccaaaaaggacgaagcattttatcatcctcgtctgattatttatgaagagg 841
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 cagagaaaaccaaaaaggacgaagcattttatcatcctcgtctgattatttatgaagagg 780

Qy     842 aagaagaggaagaggaagaggaggaggaagtagaaaatgaggaacaagattctcagaatg 901
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 aagaagaggaagaggaagaggaggaggaagtagaaaatgaggaacaagattctcagaatg 840

Qy     902 agggtagtacagatgagaagtcaagccctgccagctctcaagaagggtctccttcaggtg 961
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 agggtagtacagatgagaagtcaagccctgccagctctcaagaagggtctccttcaggtg 900

Qy     962 accagcagttttcacctaaatccaacactgaaaaatctaaaggggaattaatgtttgatg 1021
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 accagcagttttcacctaaatccaacactgaaaaatctaaaggggaattaatgtttgatg 960

Qy    1022 attcttcagattcatcaccggaaaaacaggagagaaatttaaactggaccccggccgaag 1081
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 attcttcagattcatcaccggaaaaacaggagagaaatttaaactggaccccggccgaag 1020

Qy    1082 tcccacagttagctgcagcaaaacgcaggctgcctcagggaaaggagcctgggttgatta 1141
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 tcccacagttagctgcagcaaaacgcaggctgcctcagggaaaggagcctgggttgatta 1080

Qy    1142 acttgtgtgccaatgtcccacccgtcccaggtaacattttgccccctgaggtccggggta 1201
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 acttgtgtgccaatgtcccacccgtcccaggtaacattttgccccctgaggtccggggta 1140

Qy    1202 atttaatggctgctggacaaaacctccaaagttctgaaagatcagaaatgatagctacct 1261
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 atttaatggctgctggacaaaacctccaaagttctgaaagatcagaaatgatagctacct 1200

Qy    1262 ggagtccagctgtacggacactgaggaatattactaataatgctgacattcagcagatga 1321
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 ggagtccagctgtacggacactgaggaatattactaataatgctgacattcagcagatga 1260

Qy    1322 accggccatcaaatgtagcacatatcttacagactctttcagcacctacgaaaaatttag 1381
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 accggccatcaaatgtagcacatatcttacagactctttcagcacctacgaaaaatttag 1320

Qy    1382 aacagcaggtgaatcacagccagcagggacatacaaatgccaatgcagtgctgtttagcc 1441
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 aacagcaggtgaatcacagccagcagggacatacaaatgccaatgcagtgctgtttagcc 1380

Qy    1442 aagtgaaagtgactccagagacacacatgctacagcagcagcagcaggcccagcagcagc 1501
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 aagtgaaagtgactccagagacacacatgctacagcagcagcagcaggcccagcagcagc 1440

Qy    1502 agcagcagcacccggttttacaccttcagccccagcagataatgcagctccagcagcagc 1561
```

```
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1441 agcagcagcacccggtttacaccttcagccccagcagataatgcagctccagcagcagc 1500

Qy     1562 agcagcagcagatctctcagcaaccttaccccagcagccgccgcatccattttcacagc 1621
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1501 agcagcagcagatctctcagcaaccttaccccagcagccgccgcatccattttcacagc 1560

Qy     1622 agcagcagcagcaacagcagccaccaccatcgcctcagcagcatcagcttttggacatg 1681
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1561 agcagcagcagcaacagcagccaccaccatcgcctcagcagcatcagcttttggacatg 1620

Qy     1682 atccagcagtggagattccagaagaaggcttcttattgggatgtgtgtttgcaattgcgg 1741
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1621 atccagcagtggagattccagaagaaggcttcttattgggatgtgtgtttgcaattgcgg 1680

Qy     1742 attatccagagcagatgtctgataagcaactgctggccacctggaaaaggataatccagg 1801
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1681 attatccagagcagatgtctgataagcaactgctggccacctggaaaaggataatccagg 1740

Qy     1802 cacatggcggcactgttgaccccaccttcacgagtcgatgcacgcaccttctctgtgaga 1861
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1741 cacatggcggcactgttgaccccaccttcacgagtcgatgcacgcaccttctctgtgaga 1800

Qy     1862 gtcaagtcagcagcgcgtatgcacaggcaataagagaaagaaagagatgtgttactgcac 1921
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1801 gtcaagtcagcagcgcgtatgcacaggcaataagagaaagaaagagatgtgttactgcac 1860

Qy     1922 actggttaaacacagtcttaaaaaaaaaaaaaatggtaccgccgcaccgagcccttcact 1981
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1861 actggttaaacacagtcttaaaaaaaaaaaaaatggtaccgccgcaccgagcccttcact 1920

Qy     1982 tcccagtggccttcccaccaggaggaaagccatgttcacagcatattatttctgtgactg 2041
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1921 tcccagtggccttcccaccaggaggaaagccatgttcacagcatattatttctgtgactg 1980

Qy     2042 gatttgttgatagtgacagagatgacctaaaattaatggcttatttggcaggtgccaaat 2101
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1981 gatttgttgatagtgacagagatgacctaaaattaatggcttatttggcaggtgccaaat 2040

Qy     2102 atacgggttatctatgccgcagcaacacagtcctcatctgtaaagaaccaactggtttaa 2161
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2041 atacgggttatctatgccgcagcaacacagtcctcatctgtaaagaaccaactggtttaa 2100

Qy     2162 agtatgaaaaagccaaagagtggaggataccctgtgtcaacgcccagtggcttggcgaca 2221
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2101 agtatgaaaaagccaaagagtggaggataccctgtgtcaacgcccagtggcttggcgaca 2160

Qy     2222 ttcttctgggaaactttgaggcactgaggcagattcagtatagtcgctacacggcattca 2281
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2161 ttcttctgggaaactttgaggcactgaggcagattcagtatagtcgctacacggcattca 2220

Qy     2282 gtctgcaggatccatttgcccctacccagcatttagttttaaatcttttagatgcttgga 2341
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2221 gtctgcaggatccatttgcccctacccagcatttagttttaaatcttttagatgcttgga 2280

Qy     2342 gagttcccttaaaagtgtctgcagagttgttgatgagtataagactacctcccaaactga 2401
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2281 gagttcccttaaaagtgtctgcagagttgttgatgagtataagactacctcccaaactga 2340

Qy     2402 aacagaatgaagtagctaatgtccagccttcttcccaaagagccagaattgaagacgtac 2461
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2341 aacagaatgaagtagctaatgtccagccttcttcccaaagagccagaattgaagacgtac 2400

Qy     2462 cacctcccactaaaaagctaactccagaattgaccccttttgtgcttttcactggattcg 2521
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2401 cacctcccactaaaaagctaactccagaattgaccccttttgtgcttttcactggattcg 2460

Qy     2522 agcctgtccaggttcaacagtatattaagaagctctacattcttggtggagaggttgcgg 2581
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2461 agcctgtccaggttcaacagtatattaagaagctctacattcttggtggagaggttgcgg 2520

Qy     2582 agtctgcacagaagtgcacacacctcattgccagcaaagtgactcgcaccgtgaagttcc 2641
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2521 agtctgcacagaagtgcacacacctcattgccagcaaagtgactcgcaccgtgaagttcc 2580

Qy     2642 tgacggcgatttctgtcgtgaagcacatagtgacgccagagtggctggaagaatgcttca 2701
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2581 tgacggcgatttctgtcgtgaagcacatagtgacgccagagtggctggaagaatgcttca 2640

Qy     2702 ggtgtcagaagttcattgatgagcagaactacattctccgagatgctgaggcagaagtac 2761
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2641 ggtgtcagaagttcattgatgagcagaactacattctccgagatgctgaggcagaagtac 2700

Qy     2762 ttttctctttcagcttggaagaatccttaaaacgggcacacgtttctccactctttaagg 2821
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2701 ttttctctttcagcttggaagaatccttaaaacgggcacacgtttctccactctttaagg 2760

Qy     2822 caaaatatttttacatcacacctggaatctgcccaagtctttccactatgaaggcaatcg 2881
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2761 caaaatatttttacatcacacctggaatctgcccaagtctttccactatgaaggcaatcg 2820

Qy     2882 tagagtgtgcaggaggaaaggtgttatccaagcagccatctttccggaagctcatggagc 2941
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2821 tagagtgtgcaggaggaaaggtgttatccaagcagccatctttccggaagctcatggagc 2880

Qy     2942 acaagcagaactcgagtttgtcggaaataattttaatatcctgtgaaaatgaccttcatt 3001
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2881 acaagcagaactcgagtttgtcggaaataattttaatatcctgtgaaaatgaccttcatt 2940

Qy     3002 tatgccgagaatattttgccagaggcatagatgttcacaatgcagagttcgttctgactg 3061
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2941 tatgccgagaatattttgccagaggcatagatgttcacaatgcagagttcgttctgactg 3000

Qy     3062 gagtgctcactcaaacgctggactatgaatcatataagtttaactgatggcgtctaggct 3121
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3001 gagtgctcactcaaacgctggactatgaatcatataagtttaactgatggcgtctaggct 3060

Qy     3122 gccgtgcatgtcgactcctgcggtgcggggctggctgtctggctggcgaggagctgctgc 3181
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3061 gccgtgcatgtcgactcctgcggtgcggggctggctgtctggctggcgaggagctgctgc 3120

Qy     3182 gcttccttcacatgctcttgttttccagctgctttcctggggggatcagactgtgaagcag 3241
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3121 gcttccttcacatgctcttgttttccagctgctttcctggggggatcagactgtgaagcag 3180

Qy     3242 gaagacagatataataaatatactgcatctttttaagatgtgcaattttattctgaggaa 3301
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3181 gaagacagatataataaatatactgcatctttttaagatgtgcaattttattctgaggaa 3240

Qy     3302 acataaattatgttttgtattatatgactttaagagcccacattaggttttatgattcat 3361
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3241 acataaattatgttttgtattatatgactttaagagcccacattaggttttatgattcat 3300

Qy     3362 ttgccaggtttttaaatgttttcacaaaactgttacgggacttcaactagaaataaaatg 3421
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3301 ttgccaggtttttaaatgttttcacaaaactgttacgggacttcaactagaaataaaatg 3360

Qy     3422 gtgtaaataaagaccttgctatctctaaattatggatgttaaagatttgaaatgttttgt 3481
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3361 gtgtaaataaagaccttgctatctctaaattatggatgttaaagatttgaaatgttttgt 3420

Qy     3482 actttgattattttatttcttatactctgttttcttttatattgatatcttgcccacat 3541
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3421 actttgattattttatttcttatactctgttttcttttatattgatatcttgcccacat 3480

Qy     3542 tttaaataaatgtacttttgaacttagaaaaaaaaaaaa 3580
            |||||||||||||||||||||||||||||||||||||||
Db     3481 tttaaataaatgtacttttgaacttagaaaaaaaaaaaa 3519
```

```
RESULT   2
AAC27011
ID   AAC27011 standard; cDNA; 335 BP.
XX
AC   AAC27011;
XX
```

```
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 31086.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 31086; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 335 BP; 118 A; 59 C; 84 G; 71 T; 3 other;

  Query Match             9.3%;  Score 331.2;  DB 21;  Length 335;
  Best Local Similarity   98.8%;  Pred. No. 7.6e-67;
  Matches  331;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy     655 cagctaaccctcaataagaaatgcacgcatttgattgttccagagccaaaggggggaaa 714
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db       1 cagctaaccctcaataagaaatgcacgcatttgattgttccagagccaaaggggagaaa 60

Qy     715 tacgaatgtgctttaaagcgagcaagtattaaaattgtgactcctgactgggttctggat 774
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 tacgaatgtgctttaaagcgagcaagtattaaaattgtgactcctgactgggttctggat 120

Qy     775 tgcgtatcagagaaaaccaaaaaggacgaagcattttatcatcctcgtctgattatttat 834
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ngcgtatcagagaaaaccaaaaaggacgaagcattttatcatcctcgtctgattatttat 180

Qy     835 gaagaggaagaagaggaagaggaagaggaggaggaagtagaaaatgaggaacaagattct 894
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 gaagaggaagaagaggaagaggaagaggaggaggaagtagaaaatgaggaacaagattct 240

Qy     895 cagaatgagggtagtacagatgagaagtcaagccctgccagctctcaagaagggtctcct 954
           ||||||||||||||||||||||||||||||||||||::||||||||||||||||||||||
Db     241 cagaatgagggtagtacagatgagaagtcaagccctkvcagctctcaagaagggtctcct 300

Qy     955 tcaggtgaccagcagttttcacctaaatccaacac 989
           |||||||||||||||||||||||||||||||||||
Db     301 tcaggtgaccagcagttttcacctaaatccaacac 335


RESULT   3
AAH70661
ID   AAH70661 standard; cDNA; 690 BP.
XX
AC   AAH70661;
XX
DT   19-SEP-2001  (first entry)
XX
DE   Human cervical cancer marker nucleic acid 1935.
XX
KW   Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO200142467-A2.
XX
PD   14-JUN-2001.
XX
PF   08-DEC-2000; 2000WO-US33312.
XX
PR   08-DEC-1999;   99US-0169681.
PR   21-DEC-1999;   99US-0171350.
PR   14-MAR-2000; 2000US-0189315.
PR   12-MAY-2000; 2000US-0203791.
PR   09-JUN-2000; 2000US-0210600.
PR   21-JUL-2000; 2000US-0220114.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R,  Deeds J,  Berger A,  Zhao X;
XX
DR   WPI; 2001-375006/39.
XX
PT   New isolated nucleic acid for diagnosing and treating cervical cancer
PT   and for assessing and detecting compounds for treating the cancer -
XX
PS   Claim 1; Page 415; 1051pp; English.
XX
CC   The invention relates to novel genes (AAH68727-AAH73383) associated with
CC   cervical cancer with cytostatic activity. The nucleic acids and encoded
CC   polypeptides are useful: to assess if a patient is afflicted with
CC   cervical cancer or has a pre-malignant condition; to monitor the
CC   progression of cervical cancer or a premalignant condition in a patient;
CC   and to select and/or assess the efficacy of a compound or therapy for
CC   inhibiting cervical cancer in a patient. The nucleic acids may also be
CC   useful for gene therapy.
XX
SQ   Sequence 690 BP; 218 A; 174 C; 172 G; 124 T; 2 other;

  Query Match             5.7%;  Score 204.6;  DB 22;  Length 690;
  Best Local Similarity   95.9%;  Pred. No. 2.1e-37;
  Matches  210;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy    1073 cggccgaagtcccacagttagctgcagcaaaacgcaggctgcctcagggaaaggagcctg 1132
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      23 cggccgaagtcccacagttagctgcagcaaaacgcaggctgcctcagggaaaggagcctg 82

Qy    1133 ggttgattaacttgtgtgccaatgtcccacccgtcccaggtaacattttgccccctgagg 1192
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db      83 ggttgattaacttgtgtgtcaatgtcccacccgtcccaggtaacattttgccccctgagg 142

Qy    1193 tccggggtaatttaatggctgctggacaaaacctccaaagttctgaaagatcagaaatga 1252
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     143 tccggggtaatttaatggctgctggacaaaacctccaaagttctgaaagatcagaaatga 202

Qy    1253 tagctacctggagtccagctgtacggacactgaggaata 1291
           |||||||||||||||||||||||||    || | | |
Db     203 tagctacctggagtccagctgtacggcacttggcgtaaa 241
```

```
RESULT   4
AAH71865
ID   AAH71865 standard; cDNA; 504 BP.
XX
AC   AAH71865;
XX
DT   19-SEP-2001  (first entry)
XX
DE   Human cervical cancer marker nucleic acid 3139.
XX
KW   Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO200142467-A2.
XX
PD   14-JUN-2001.
XX
PF   08-DEC-2000; 2000WO-US33312.
XX
PR   08-DEC-1999;   99US-0169681.
PR   21-DEC-1999;   99US-0171350.
PR   14-MAR-2000; 2000US-0189315.
PR   12-MAY-2000; 2000US-0203791.
PR   09-JUN-2000; 2000US-0210600.
PR   21-JUL-2000; 2000US-0220114.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R,  Deeds J,  Berger A,  Zhao X;
XX
DR   WPI; 2001-375006/39.
XX
PT   New isolated nucleic acid for diagnosing and treating cervical cancer
PT   and for assessing and detecting compounds for treating the cancer -
XX
PS   Claim 1; Page 622; 1051pp; English.
XX
CC   The invention relates to novel genes (AAH68727-AAH73383) associated with
CC   cervical cancer with cytostatic activity. The nucleic acids and encoded
CC   polypeptides are useful: to assess if a patient is afflicted with
CC   cervical cancer or has a pre-malignant condition; to monitor the
CC   progression of cervical cancer or a premalignant condition in a patient;
CC   and to select and/or assess the efficacy of a compound or therapy for
CC   inhibiting cervical cancer in a patient. The nucleic acids may also be
CC   useful for gene therapy.
XX
SQ   Sequence 504 BP; 147 A; 135 C; 121 G; 101 T; 0 other;


  Query Match             5.2%;  Score 186.8;  DB 22;  Length 504;
  Best Local Similarity   94.0%;  Pred. No. 2.4e-33;
  Matches  205;  Conservative    0;  Mismatches   12;  Indels    1;  Gaps    1;

Qy   1075 gccgaagtcccacagttagctgcagcaaaacgcaggctgcctcagggaaaggagcctggg 1134
           | |  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 ggccgagtcccacagttagctgcagcaaaacgcaggctgcctcagggaaaggagcctggg 80

Qy   1135 ttgattaacttgtgtgccaatgtcccacccgtcccaggtaacattttgccccctgaggtc 1194
           ||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db     81 ttgattaacttgtgtgtcaatgtcccacccgtcccaggtaacattttgccccctgaggtc 140

Qy   1195 cggggtaatttaatggctgctggacaaaa:::ccaaagttc-tgaaagatcagaaatgat 1253
           |||||||||||||||||||||||||||!|||||||||||| ||||||||||||||||||
Db    141 cggggtaatttaatggctgctggacaaaacctccaaagttcttgaaagatcagaaatgat 200

Qy   1254 agctacctggagtccagctgtacggacactgaggaata 1291
           |||||||||||||||||||||||||    !!  | | |
Db    201 agctacctggagtccagctgtacggcacttggcgtaaa 238


RESULT   5
AAH68998
ID   AAH68998 standard; cDNA; 579 BP.
XX
AC   AAH68998;
XX
DT   19-SEP-2001  (first entry)
XX
DE   Human cervical cancer marker nucleic acid 272.
XX
KW   Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO200142467-A2.
XX
PD   14-JUN-2001.
XX
PF   08-DEC-2000; 2000WO-US33312.
XX
PR   08-DEC-1999;   99US-0169681.
PR   21-DEC-1999;   99US-0171350.
PR   14-MAR-2000; 2000US-0189315.
PR   12-MAY-2000; 2000US-0203791.
PR   09-JUN-2000; 2000US-0210600.
PR   21-JUL-2000; 2000US-0220114.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R,  Deeds J,  Berger A,  Zhao X;
XX
DR   WPI; 2001-375006/39.
XX
PT   New isolated nucleic acid for diagnosing and treating cervical cancer
PT   and for assessing and detecting compounds for treating the cancer -
XX
PS   Claim 1; Page 150-151; 1051pp; English.
XX
CC   The invention relates to novel genes (AAH68727-AAH73383) associated with
CC   cervical cancer with cytostatic activity. The nucleic acids and encoded
CC   polypeptides are useful: to assess if a patient is afflicted with
CC   cervical cancer or has a pre-malignant condition; to monitor the
CC   progression of cervical cancer or a premalignant condition in a patient;
CC   and to select and/or assess the efficacy of a compound or therapy for
CC   inhibiting cervical cancer in a patient. The nucleic acids may also be
CC   useful for gene therapy.
XX
SQ   Sequence 579 BP; 169 A; 152 C; 145 G; 110 T; 3 other;


  Query Match             5.2%;  Score 186.8;  DB 22;  Length 579;
  Best Local Similarity   94.0%;  Pred. No. 2.5e-33;
  Matches  205;  Conservative    0;  Mismatches   12;  Indels    1;  Gaps    1;

Qy   1075 gccgaagtcccacagttagctgcagcaaaacgcaggctgcctcagggaaaggagcctggg 1134
           | |  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 ggccgagtcccacagttagctgcagcaaaacgcaggctgcctcagggaaaggagcctggg 80

Qy   1135 ttgattaacttgtgtgccaatgtcccacccgtcccaggtaacattttgccccctgaggtc 1194
           ||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db     81 ttgattaacttgtgtgtcaatgtcccacccgtcccaggtaacattttgccccctgaggtc 140

Qy   1195 cggggtaatttaatggctgctggacaaaacctccaaagttc-tgaaagatcagaaatgat 1253
           |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    141 cggggtaatttaatggctgctggacaaaacctccaaagttcttgaaagatcagaaatgat 200

Qy   1254 agctacctggagtccagctgtacggacactgaggaata 1291
           |||||||||||||||||||||||||    !!  | | |
Db    201 agctacctggagtccagctgtacggcacttggcgtaaa 238


RESULT   6
AAI85135
ID   AAI85135 standard; cDNA; 467 BP.
```

```
XX
AC   AAI85135;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 5195.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation; ss.
XX
OS   Homo sapiens.
XX
PN   WO200164835-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04927.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   P-PSDB; AAO05204.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing
PT   diagnosing and treating e.g. leukaemia, inflammation and immune
PT   disorders -
XX
PS   Claim 1; SEQ ID NO 5195; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC   the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 467 BP; 122 A; 173 C; 107 G; 61 T; 4 other;


  Query Match             4.5%;  Score 162;  DB 22;  Length 467;
  Best Local Similarity   94.4%;  Pred. No. 1.3e-27;
  Matches  168;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy   1472 tacagcagcagcagcaggcccagcagcagcagcagcagcacccggttttacaccttcagc 1531
           ||||||||||||||| ||||||| ||||||||||||||||||||||||||||||||||||
Db     40 tacagcagcagcagcgggcccaggagcagcagcagcagcacccggttttacaccttcagc 99

Qy   1532 cccagcagataatgcagctccagcagcagcagcagcagcagatctctcagcaaccttacc 1591
           |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    100 cccagcagataatgcagctccagcagcagcagcagcggcagatctctcagcaaccttacc 159

Qy   1592 cccagcagccgccgcatccatttttcacagcagcagcagcagcaacagcagccaccacc 1649
           |||||||||||||||||||||||||||||| |||||||||| ||||| | |  ||
Db    160 cccagcagccgccgcatccatttttcacagcaacagcagcagcagcagcaagcccatcc 217


RESULT   7
AAX11992
```

```
ID   AAX11992 standard; DNA; 150 BP.
XX
AC   AAX11992;
XX
DT   30-MAR-1999  (first entry)
XX
DE   Human biallelic polymorphic DNA fragment WI-18169.
XX
KW   Polymorphism; biallelic; human; forensic; paternity testing; disease;
KW   detection; phenotypic typing; characteristic; infection; hereditary;
KW   autoimmune disease; cancer; inflammation; drug; therapy; medicament;
KW   treatment; marker; ss.
XX
OS   Homo sapiens.
XX
PN   WO9820165-A2.
XX
PD   14-MAY-1998.
XX
PF   05-NOV-1997;   97WO-US20313.
XX
PR   06-NOV-1996;   96US-0030455.
XX
PA   (WHED ) WHITEHEAD INST BIOMEDICAL RES.
XX
PI   Hudson T, Lander ES, Wang D;
XX
DR   WPI; 1998-286974/25.
XX
PT   New isolated nucleic acid segments from the human genome - used for
PT   determining polymorphic forms for use in e.g. forensics, paternity
PT   testing or phenotypic typing for disease
XX
PS   Claim 1; Page 210; 310pp; English.
XX
CC   AAX10269-X12937 are human DNA fragments which contain biallelic
CC   polymorphic markers which have been isolated using the primers
CC   represented in AAX09121-X10268. The base occupying the polymorphic site
CC   is indicated by the appropriate IUPAC-IUB ambiguity code. These fragments
CC   can be used in methods for determining polymorphic forms in an individual
CC   for use in e.g. forensics, paternity testing or for phenotypic typing for
CC   diseases such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan
CC   syndrome, muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease,
CC   familial hypercholesterolemia, polycystic kidney disease, hereditary
CC   spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary
CC   haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos
CC   syndrome, osteogenesis imperfecta, acute intermittent porphyria,
CC   autoimmune diseases, inflammation, cancer, diseases of the nervous
CC   system, infection by pathogenic microorganisms, and characteristics such
CC   as longevity, appearance (e.g. baldness, obesity), strength, speed,
CC   endurance, fertility, and susceptibility or receptivity to particular
CC   drugs or therapeutic treatments. The isolated polymorphic nucleic acid
CC   segments can also be used to produce medicaments for the treatment or
CC   prophylaxis of such diseases.
XX
SQ   Sequence 150 BP; 43 A; 32 C; 40 G; 34 T; 1 other;


  Query Match             3.8%;  Score 135.6;  DB 19;  Length 150;
  Best Local Similarity   99.3%;  Pred. No. 9.8e-22;
  Matches  135;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   2820 ggcaaaatatttttacatcacacctggaatctgcccaagtctttccactatgaaggcaat 2879
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ggcaaaatatttttacatcacacctggaatctgcccaagtctttccactatgaaggcaat 60

Qy   2880 cgtagagtgtgcaggaggaaaggtgttatccaagcagccatctttccggaagctcatgga 2939
           |||||||||||||||||||||||||||||||||||||||||||||||||||||:||||
Db     61 cgtagagtgtgcaggaggaaaggtgttatccaagcagccatctttccggaagctcrtgga 120

Qy   2940 gcacaagcagaactcg 2955
           ||||||||||||||||
Db    121 gcacaagcagaactcg 136
```

```
RESULT   8
AAS44894/c
ID   AAS44894 standard; DNA; 344 BP.
XX
AC   AAS44894;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Human contig polynucleotide sequence #147.
XX
KW   Mammal; human; rhesus monkey; baker's yeast; fission yeast; Norway rat;
KW   mouse; Chinese hamster; African clawed frog; fruit fly; dog; leukaemia;
KW   cancer; lymphoma; neuroblastoma; autoimmune disorder; cell proliferation;
KW   nervous system disorder; inflammatory disorder; cell differentiation; ds;
KW   angiogenesis; stem cell growth factor; activin; inhibin; cartilage; burn;
KW   genetic disorder; bone regeneration; tendon; ligament; tissue repair;
KW   cytostatic; antirheumatic; antiarthritic; vulnerary; antiinflammatory;
KW   antibacterial; immunosuppressive; vasotropic; antiparkinsonian;
KW   neuroprotective; osteopathic; antidiabetic; antiasthmatic; antiallergic;
KW   immunostimulant; analgesic; gene therapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO200164834-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04926.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
PR   17-JUN-2000; 2000US-0597707.
PR   14-JUL-2000; 2000US-0616807.
PR   19-SEP-2000; 2000US-0664641.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Zhou P,  Asundi V,  Zhang J,  Zhao QA,  Ren F;
PI   Xue AJ,  Yang Y,  Wehrman T,  Wang J,  Ma Y,  Wang D,  Chen R,  Xu C;
PI   Drmanac R;
XX
DR   WPI; 2001-589862/66.
DR   P-PSDB; AAU27994.
XX
PT   Novel polypeptides and nucleic acids obtained from cDNA libraries
PT   prepared from various human tissues, for diagnosis, treatment of
PT   cancer, neurological, inflammatory disorders and for use in arrays for
PT   detection -
XX
PS   Claim 1; SEQ ID No 491; 153pp; English.
XX
CC   Sequences AAS44576-AAS44919 represent full-length polynucleotides and
CC   contig polynucleotides encoding polypeptides of the invention. The DNA
CC   and protein sequences are useful for the treatment, diagnosis and
CC   prevention of various types of disorder in a mammalian subject such as a
CC   human, dog, monkey, mouse, hamster or rat. The disorders include cancers
CC   such as leukaemia, lymphoma and neuroblastoma, autoimmune disorders such
CC   as multiple sclerosis, connective tissue disease, rheumatoid arthritis,
CC   diabetes mellitus, allergic rhinitis, asthma and eczema, nervous system
CC   disorders such as Parkinson's disease, Alzheimer's disease, Huntington's
CC   chorea, amyotrophic lateral sclerosis, spinal muscular atrophy and
CC   Wernicke disease, inflammatory disorders such as nephritis, Crohn's
CC   disease, ischaemia-reperfusion injury, shock, sepsis and inflammatory
CC   bowel disease. The sequences exhibit activity relating to angiogenesis,
CC   cell proliferation, cell differentiation, stem cell growth factor,
CC   activin or inhibin. Therefore, they can be used to manipulate stem cells
CC   in culture to give rise to neuroepithelial cells that can be used to
CC   augment or replace cells damaged by illness, accidental damage or genetic
CC   disorders. The sequences may also be used for regeneration of bone,
CC   cartilage, tendons and ligaments and in tissue repair and burn healing.
CC   Note: Some sequences for this patent did not form part of the printed
CC   specification, but were obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 344 BP; 97 A; 83 C; 79 G; 85 T; 0 other;


  Query Match             3.5%;  Score 125;  DB 22;  Length 344;
  Best Local Similarity   100.0%;  Pred. No. 4e-19;
  Matches  125;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    2831 tttacatcacacctggaatctgcccaagtctttccactatgaaggcaatcgtagagtgtg 2890
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     344 TTTACATCACACCTGGAATCTGCCCAAGTCTTTCCACTATGAAGGCAATCGTAGAGTGTG 285

Qy    2891 caggaggaaaggtgttatccaagcagccatctttccggaagctcatggagcacaagcaga 2950
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     284 CAGGAGGAAAGGTGTTATCCAAGCAGCCATCTTTCCGGAAGCTCATGGAGCACAAGCAGA 225

Qy    2951 actcg 2955
           |||||
Db     224 ACTCG 220


RESULT   9
AAX89891
ID   AAX89891 standard; DNA; 397 BP.
XX
AC   AAX89891;
XX
DT   05-NOV-1999  (first entry)
XX
DE   Spinocerebellar ataxia type III (SCAIII) gene fragment.
XX
KW   Spinocerebellar ataxia type III; SCAIII; reverse dot hybridisation;
KW   PCR-microplate hybridisation; PCR-MPH; trinucleotide repeat; TNR;
KW   SCAIII syndrome; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   repeat_region   137..355
FT                   /*tag= a
FT   repeat_unit     137..139
FT                   /*tag= b
FT                   /note= "trinucleotide repeat"
XX
PN   WO9943852-A1.
XX
PD   02-SEP-1999.
XX
PF   18-FEB-1999;   99WO-KR00078
XX
PR   26-FEB-1998;   98KR-0006278.
XX
PA   (JIND/) JIN D K.
PA   (SMSU ) SAMSUNG FINE CHEM CO LTD.
XX
PI   Jin DK;
XX
DR   WPI; 1999-527634/44.
XX
PT   Diagnosis of spinocerebellar ataxia type III (SCA III) syndrome
PT   using techniques which ensure highly accurate diagnosis
XX
PS   Claim 1; Page 12-13; 28pp; English.
XX
CC   The invention relates to the diagnosis of spinocerebellar ataxia type
CC   III (SCAIII) syndrome using reverse dot hybridisation or PCR-microplate
CC   hybridisation (PCR-MPH). The method comprises attaching a potion of the
CC   SCAIII gene containing 73 copies of the trinucleotide (CAG) repeat unit
CC   (the present sequence) to a substrate, and hybridising with amplified
CC   testee genomic DNA containing copies of the trinucleotide units, PCR
```

```
CC   amplified with labeled primers (AAX89889-90). The new method is useful
CC   for diagnosis of SCAIII syndrome, and for determining the severity of
CC   the disease. The present sequence represents the SCAIII gene fragment
CC   containing 73 trinucleotide (TNR) repeats.
XX
SQ   Sequence 397 BP; 124 A; 104 C; 104 G; 65 T; 0 other;


  Query Match             3.1%;  Score 110.6;  DB 20;  Length 397;
  Best Local Similarity   63.2%;  Pred. No. 9.3e-16;
  Matches  170;  Conservative    0;  Mismatches   99;  Indels    0;  Gaps    0;

Qy   1435 tttagccaagtgaaagtgactccagagacacacatgctacagcagcagcagcaggcccag 1494
           || || ||   | ||  |   | | ||   ||   ||  ||||||||||||||||   |||
Db    113 ttcagacagcagcaaaagcagcaacagcagcagcagcagcagcagcagcagcagcagcag 172

Qy   1495 cagcagcagcagcagcacccggttttacaccttcagccccagcagataatgcagctccag 1554
           ||||||||||||||||| | |      || |  ||||  ||||||     |||||  |||
Db    173 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 232

Qy   1555 cagcagcagcagcagcagatctctcagcaaccttacccccagcagccgccgcatccattt 1614
           ||||||||||||||||||      ||||| |   | |  ||||||| || ||| |
Db    233 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 292

Qy   1615 tcacagcagcagcagcagcaacagcagccaccaccatcgcctcagcagcatcagctttttt 1674
             |||||||||||||||||| |||||||  |  |    ||  |||||||| ||||
Db    293 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 352

Qy   1675 ggacatgatccagcagtggagattccaga 1703
             |  || | | |||   ||| | || |
Db    353 cagcgggacctatcaggacagagttcaca 381


RESULT  10
AAC84677
ID   AAC84677 standard; DNA; 5120 BP.
XX
AC   AAC84677;
XX
DT   20-APR-2001  (first entry)
XX
DE   DNA sequence of human PPP2R2B.
XX
KW   Microsatellite marker; human; chromosome 5q31-33; D5S402; WI-6763;
KW   CAG trinucleotide repeat; spinocerebellar ataxia 12; SCA12; PPP2RBbeta;
KW   brain; protein phosphatase PP2A; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   repeat_region    2088..2366
FT                    /*tag=  a
FT                    /rpt_type= "CAG"
FT   mRNA             2473..3093
FT                    /*tag=  b
FT   5'UTR            2473..3023
FT                    /*tag=  c
FT   CDS              3024..3093
FT                    /*tag=  d
FT                    /product= ""PPP2R2B""
FT                    /note= "protein _id= AAF74024.1"
XX
PN   WO200078943-A2.
XX
PD   28-DEC-2000.
XX
PF   16-JUN-2000; 2000WO-US40213.
XX
PR   18-JUN-1999;   99US-0140176.
XX
PA   (UYJO ) UNIV JOHNS HOPKINS.
XX
PI   Margolis RL,  Ross CA,  Holmes SE;
XX
DR   WPI; 2001-102718/11.
DR   P-PSDB; AAB48340.
XX
PT   New polynucleotide for use in diagnosing spinocerebellar ataxia 12,
PT   comprises a microsatellite marker having a variable number of CAG
PT   trinucleotide repeats located on human chromosome 5q31-33
XX
PS   Disclosure; Page 21-23; 23pp; English.
XX
CC   The invention relates to a microsatellite marker located on human
CC   chromosome 5q31-33 between markers D5S402 and WI-6763, comprising a
CC   variable number of CAG trinucleotide repeats. The marker is useful for
CC   diagnosing spinocerebellar ataxia 12 (SCA12) by determining the number of
CC   CAG trinucleotide repeats in a 5'-untranslated region of an allele of
CC   PPP2RBbeta, a brain specific regulatory subunit of protein phosphatase
CC   PP2A, where if a number greater than 40 is determined, an expansion which
CC   is associated with SCA12 is indicated. The presence of expanded
CC   trinucleotide repeats on chromosome 5q31-33 is useful for predicting or
CC   diagnosing SCA12. Also provided are primers useful for amplifying the
CC   microsatellite marker associated with SCA12 and for diagnosing SCA12 and
CC   a probe useful for detecting expansions in a microsatellite marker
CC   located on human chromosome 5q31-33. The present sequence represents
CC   a PPP2R2B DNA sequence.
XX
SQ   Sequence 5120 BP; 1336 A; 1267 C; 1268 G; 1249 T; 0 other;


  Query Match             2.9%;  Score 105.6;  DB 22;  Length 5120;
  Best Local Similarity   62.5%;  Pred. No. 4.2e-14;
  Matches  165;  Conservative    0;  Mismatches   99;  Indels    0;  Gaps    0;

Qy   1406 agggacatacaaatgccaatgcagtgctgtttagccaagtgaaagtgactccagagacac 1465
           || | | |     |||  | || ||   |   ||||  |         | ||||   |
Db   2017 agagtcgtg ggctgctgacgcggttgggaggagcctcgcctttaatgcaccagccgcct 2076

Qy   1466 acatgctacagcagcagcagcaggcccagcagcagcagcagcagcacccggttttacacc 1525
            ||  || | ||||||||||||||   ||||||||||||||||||||| | |      || |
Db   2077 ccagcctcctgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2136

Qy   1526 ttcagccccagcagataatgcagctccagcagcagcagcagcagatctctcagcaac 1585
            ||||  ||||||     |||||  ||||||||||||||||||||||      |||||
Db   2137 agcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2196

Qy   1586 cttacccccagcagccgccgcatccattttcacagcagcagcagcagcaacagcagccac 1645
            | |  |||||||  || ||| |        ||||||||||||||||| |||||||  |
Db   2197 agcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2256

Qy   1646 caccatcgcctcagcagcatcagc 1669
            |    ||  |||||||| ||||
Db   2257 agcagcagcagcagcagcagcagc 2280


RESULT  11
AAA59242/c
ID   AAA59242 standard; cDNA; 1037 BP.
XX
AC   AAA59242;
XX
DT   07-NOV-2000  (first entry)
XX
DE   Exons E, C and A of the SCA8 repeat region.
XX
KW   Repeat region; spinocerebellar ataxia type 8; SCA8; long arm;
KW   chromosome 13; polymorphic CTA repeat; CTG repeat; ss.
XX
OS   Homo sapiens.
XX
PN   CA2283758-A1.
XX
PD   28-APR-2000.
XX
PF   27-SEP-1999;   99CA-2283758.
```

XX
PR 28-OCT-1998; 98US-0181585.
XX
PA (MINU) UNIV MINNESOTA.
XX
PI Benzow KA, Moseley-alldredge ML, Ranum LPW, Koob MD;
XX
DR WPI; 2000-491456/44.
XX
PT Novel nucleic acid molecule containing repeat region of spinocerebellar
PT ataxia type 8 coding sequence, useful as primer or probe for diagnosing
PT or determining risk for developing spinocerebellar ataxia -
XX
PS Claim 25; Fig 7C; 68pp; English.
XX
CC The present sequence represents a nucleic acid molecule containing a
CC repeat region of a spinocerebellar ataxia type 8 (SCA8) coding
CC sequence which is located within the long arm of chromosome 13. The
CC SCA8 coding sequence contains polymorphic CTA repeats and CTG repeats.
CC Generally, the CTA repeat is unstable, and can vary between individuals
CC in different families (the number of CTA repeats in the region does not
CC vary between individuals of a family). Fragments of the nucleic acid
CC molecule are useful as primers and probes for determining whether an
CC individual has or is at risk of developing spinocerebellar ataxia 8.
XX
SQ Sequence 1037 BP; 229 A; 261 C; 242 G; 305 T; 0 other;

Query Match 2.9%; Score 105.2; DB 21; Length 1037;
Best Local Similarity 69.4%; Pred. No. 2.5e-14;
Matches 143; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

Qy 1464 acacatgctacagcagcagcagcaggcccagcagcagcagcagcagcacccggttttaca 1523
Db 920 AAAAATGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 861

Qy 1524 ccttcagccccagcagataatgcagctccagcagcagcagcagcagcagatctctcagca 1583
Db 860 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 801

Qy 1584 accttacccccagcagccgccgcatccatttcacagcagcagcagcagcaacagcagcc 1643
Db 800 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 741

Qy 1644 accaccatcgcctcagcagcatcagc 1669
Db 740 GCAGCAGCAGCAGCAGCAGCAGC 715

RESULT 12
AAA59240/c
ID AAA59240 standard; DNA; 1159 BP.
XX
AC AAA59240;
XX
DT 07-NOV-2000 (first entry)
XX
DE An EcoRI fragment that includes SCA8 repeat region.
XX
KW Repeat region; spinocerebellar ataxia type 8; SCA8; long arm;
KW chromosome 13; polymorphic CTA repeat; CTG repeat; ss.
XX
OS Homo sapiens.
XX
PN CA2283758-A1.
XX
PD 28-APR-2000.
XX
PF 27-SEP-1999; 99CA-2283758.
XX
PR 28-OCT-1998; 98US-0181585.
XX
PA (MINU) UNIV MINNESOTA.
XX
PI Benzow KA, Moseley-alldredge ML, Ranum LPW, Koob MD;
XX
DR WPI; 2000-491456/44.
XX
PT Novel nucleic acid molecule containing repeat region of spinocerebellar
PT ataxia type 8 coding sequence, useful as primer or probe for diagnosing
PT or determining risk for developing spinocerebellar ataxia -
XX
PS Claim 26; Fig 7A; 68pp; English.
XX
CC The present sequence represents a nucleic acid molecule containing a
CC repeat region of a spinocerebellar ataxia type 8 (SCA8) coding
CC sequence which is located within the long arm of chromosome 13. The
CC SCA8 coding sequence contains polymorphic CTA repeats and CTG repeats.
CC Generally, the CTA repeat is unstable, and can vary between individuals
CC in different families (the number of CTA repeats in the region does not
CC vary between individuals of a family). Fragments of the nucleic acid
CC molecule are useful as primers and probes for determining whether an
CC individual has or is at risk of developing spinocerebellar ataxia 8.
XX
SQ Sequence 1159 BP; 263 A; 263 C; 196 G; 437 T; 0 other;

Query Match 2.9%; Score 105.2; DB 21; Length 1159;
Best Local Similarity 69.4%; Pred. No. 2.7e-14;
Matches 143; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

Qy 1464 acacatgctacagcagcagcagcaggcccagcagcagcagcagcagcacccggttttaca 1523
Db 732 AAAAATGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 673

Qy 1524 ccttcagccccagcagataatgcagctccagcagcagcagcagcagcagatctctcagca 1583
Db 672 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGGAGCAGCAGCA 613

Qy 1584 accttacccccagcagccgccgcatccattttcacagcagcagcagcagcaacagcagcc 1643
Db 612 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 553

Qy 1644 accaccatcgcctcagcagcatcagc 1669
Db 552 GCAGCAGCAGCAGCAGCAGCAGC 527

RESULT 13
AAA59241/c
ID AAA59241 standard; cDNA; 1472 BP.
XX
AC AAA59241;
XX
DT 07-NOV-2000 (first entry)
XX
DE Exons D, C, B and A of the SCA8 repeat region.
XX
KW Repeat region; spinocerebellar ataxia type 8; SCA8; long arm;
KW chromosome 13; polymorphic CTA repeat; CTG repeat; ss.
XX
OS Homo sapiens.
XX
PN CA2283758-A1.
XX
PD 28-APR-2000.
XX
PF 27-SEP-1999; 99CA-2283758.
XX
PR 28-OCT-1998; 98US-0181585.
XX
PA (MINU) UNIV MINNESOTA.
XX
PI Benzow KA, Moseley-alldredge ML, Ranum LPW, Koob MD;
XX
DR WPI; 2000-491456/44.

XX
PT Novel nucleic acid molecule containing repeat region of spinocerebellar
PT ataxia type 8 coding sequence, useful as primer or probe for diagnosing
PT or determining risk for developing spinocerebellar ataxia -
XX
PS Claim 24; Fig 7B; 68pp; English.
XX
CC The present sequence represents a nucleic acid molecule containing a
CC repeat region of a spinocerebellar ataxia type 8 (SCA8) coding
CC sequence which is located within the long arm of chromosome 13. The
CC SCA8 coding sequence contains polymorphic CTA repeats and CTG repeats.
CC Generally, the CTA repeat is unstable, and can vary between individuals
CC in different families (the number of CTA repeats in the region does not
CC vary between individuals of a family). Fragments of the nucleic acid
CC molecule are useful as primers and probes for determining whether an
CC individual has or is at risk of developing spinocerebellar ataxia 8.
XX
SQ Sequence 1472 BP; 337 A; 368 C; 378 G; 388 T; 1 other;

Query Match 2.9%; Score 105.2; DB 21; Length 1472;
Best Local Similarity 69.4%; Pred. No. 3e-14;
Matches 143; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

Qy 1464 acacatgctacagcagcagcagcaggcccagcagcagcagcagcagcacccggttttaca 1523
Db 1355 AAAAATGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 1296

Qy 1524 ccttcagccccagcagataatgcagctccagcagcagcagcagcagcagatctctcagca 1583
Db 1295 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 1236

Qy 1584 accttacccccagcagccgccgcatccattttcacagcagcagcagcagcaacagcagcc 1643
Db 1235 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 1176

Qy 1644 accaccatcgcctcagcagcatcagc 1669
Db 1175 GCAGCAGCAGCAGCAGCAGCAGCC 1150

RESULT 14
AAQ84832
ID AAQ84832 standard; DNA; 234 BP.
XX
AC AAQ84832;
XX
DT 25-SEP-1995 (first entry)
XX
DE Spinocerebellar ataxia type 1 CAG repeat region patient #2.
XX
KW Spinocerebellar ataxia type 1; SCA 1; presymptomatic diagnosis;
KW CAG repeat region; patient #2; ss.
XX
OS Homo sapiens.
XX
PN WO9501437-A.
XX
PD 12-JAN-1995.
XX
PF 29-JUN-1994; 94WO-US07336.
XX
PR 29-JUN-1993; 93US-0084365.
PR 28-JUN-1994; 94US-0267803.
XX
PA (MINU) UNIV MINNESOTA.
XX
PI Chung M, Orr HT, Zoghbi HY;
XX
DR WPI; 1995-061001/08.
XX
PT New autosomal dominant spinocerebellar ataxia type 1 nucleic acid
PT - used to develop prods. for detection or presymptomatic
PT diagnosis of a SCA1 disorder
XX
PS Disclosure; Fig 2; 111pp; English.
XX
CC AAQ84831-Q84835 show the CAG repeat regions of five individuals
CC affected with spinocerebellar ataxia type 1 (SCA 1). It is within
CC CAG repeat region (AAQ84804) that the mutations responsible for
CC SCA 1 occur. The full nucleic acid (AAQ84793) and its protein product
CC (AAR71111) can be used to develop products, for the presymptomatic
CC detection of a SCA 1 disorder.
XX
SQ Sequence 234 BP; 73 A; 80 C; 77 G; 4 T; 0 other;

Query Match 2.9%; Score 103; DB 16; Length 234;
Best Local Similarity 66.4%; Pred. No. 4.2e-14;
Matches 148; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

Qy 1465 cacatgctacagcagcagcagcaggcccagcagcagcagcagcagcacccggtttacac 1524
Db 4 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 63

Qy 1525 cttcagccccagcagataatgcagctccagcagcagcagcagcagatctctcagcaa 1584
Db 64 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 123

Qy 1585 ccttacccccagcagccgccgcatccattttcacagcagcagcagcagcaacagcagcca 1644
Db 124 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 183

Qy 1645 ccaccatcgcctcagcagcatcagctttttggacatgatccag 1687
Db 184 cagcagcagcagcagcagcagcagcacctcagcagggctccgg 226

RESULT 15
AAZ44307
ID AAZ44307 standard; DNA; 477 BP.
XX
AC AAZ44307;
XX
DT 04-APR-2000 (first entry)
XX
DE Human SCA7 genomic DNA fragment 1.
XX
KW SCA7; human; spinocerebellar ataxia type 7; SCA1; SCA2; SCA3; SCA6;
KW repeat expansion detection; RED analysis; detection; ds.
XX
OS Homo sapiens.
XX
PN CA2245310-A.
XX
PD 19-FEB-1999.
XX
PF 19-AUG-1998; 98CA-2245310.
XX
PR 19-AUG-1997; 97US-0056170.
XX
PA (MINU) UNIV MINNESOTA.
XX
PI Koob MD, Ranum LP;
XX
DR WPI; 2000-098181/09.
XX
PT Identifying individuals at risk of developing spinocerebellar ataxia
PT type 7 by analyzing trinucleotide repeat regions of spinocerebellar
PT ataxia type 7 gene -
XX
PS Claim 36; Page 41; 66pp; English.
XX
CC This invention describes a novel method for identifying individuals at
CC risk for developing spinocerebellar ataxia type 7 (SCA7). The method
CC comprises analyzing the CAG repeat region of a SCA7 gene to detect CAG

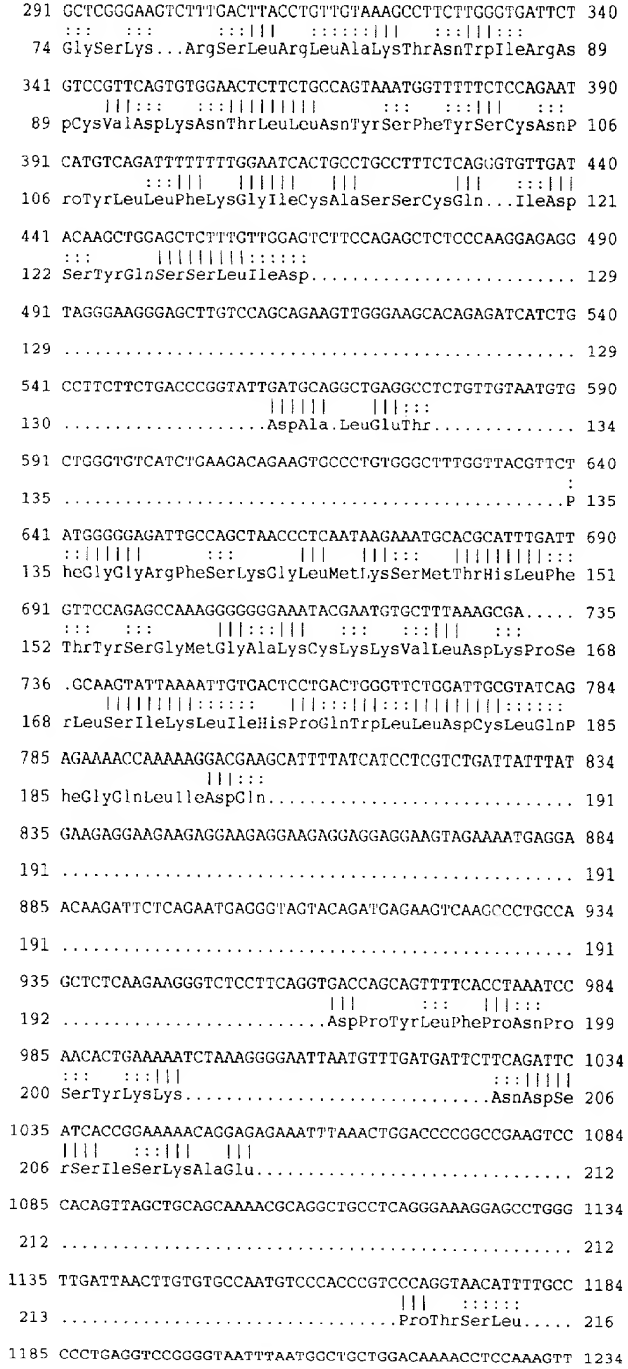
CC repeats, where individuals at risk have at least 30 CAG repeats and
CC those not at risk have less than 19 CAG repeats. The method is useful
CC for identifying individuals at risk of developing SCA7 and also those
CC at risk of developing SCA1, 2, 3 or 6. The use of genomic DNA in the
CC repeat expansion detection (RED) analysis allows isolation of any
CC potential trinucleotide repeat expansion regardless of the expression
CC pattern. Utilization of different oligonucleotides in the RED assay
CC allows any of the possible trinucleotide repeats to be detected, and
CC the cycled nature of the reaction makes it extremely sensitive. This
CC sequence represents a genomic DNA fragment of the human SCA7 gene
CC which is described in the method of the invention.
XX
SQ Sequence 477 BP; 99 A; 164 C; 172 G; 42 T; 0 other;

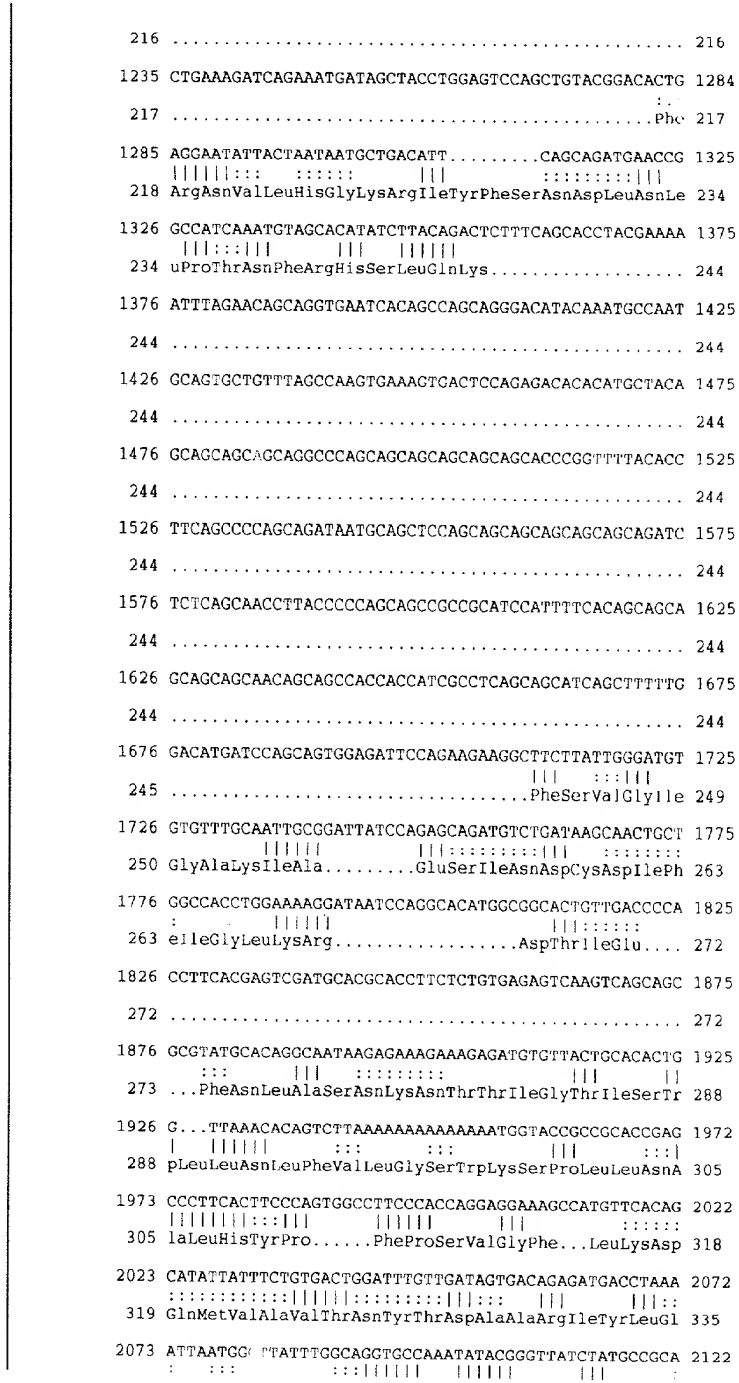
Query Match 2.8%; Score 100.2; DB 21; Length 477;
Best Local Similarity 69.9%; Pred. No. 2.6e-13;
Matches 135; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

```
Qy   1477 cagcagcagcaggcccagcagcagcagcagcagcacccggtttacaccttcagccccag 1536
          | ||||||||||   ||||||||||||||||||||| | |      || |  ||||  ||| 
Db    124 cggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 183

Qy   1537 cagataaIgcagctccagcagcagcagcagcagatctctcagcaaccttacccccag 1596
          |||     |||||  ||||||||||||||||||||      |||||  |   | |  ||| 
Db    184 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 243

Qy   1597 cagccgccgcatccattttcacagcagcagcagcagcaacagcagccaccaccatcgcct 1656
          |||| || ||| |        ||||||||||||||||| |||| ||| || ||   ||| 
Db    244 cagcagcagcagcagcagcagcagcagcagcagcagcagccgccgcctccgcagccc 303

Qy   1657 cagcagcatcagc 1669
          |||| ||| |||| 
Db    304 cagcggcagcagc 316
```

Search completed: June 4, 2002, 15:45:30
Job time: 10778 sec

```
OM of: US-09-664-641-10 to: SwissProt_40:*   out_format : pfs

Date: Jun 4, 2002  5:50 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_n2p.model -DEV=x1p
 -Q=/cgn2_1/USPTO_spool/US09664641/runat_04062002_110933_9290/app_query.fasta_1.3697
 -DB=SwissProt_40 -QFMT=fastan -SUFFIX=rsp -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09664641_@CGN1_1_66 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-664-641-10
Query length: 3580
Database: SwissProt_40:*
Database sequences: 105224
Database length: 38719550
Search time (sec): 74.270000



score_list:
Sequence              Strd Orig    ZScore   EScore Len  ! Documentation  ..
SwissProt_40:YBM5_SCHPO +  314.50  334.75  6.8e-12   878  ! Q10337 schizosaccharomyces p
SwissProt_40:SNF5_YEAST +  223.50  234.90  2.4e-06   905  ! P18480 saccharomyces cerevis
SwissProt_40:YKF4_YEAST +  213.00  225.17  1.0e-05   738  ! P35732 saccharomyces cerevis
SwissProt_40:Y083_CAEEL +  204.50  213.24  3.5e-05  1000  ! P34619 caenorhabditis elega
SwissProt_40:Y192_HUMAN +  200.00  201.82  7.1e-05  2124  ! Q93074 homo sapiens (human)
SwissProt_40:FXP2_MOUSE +  198.50  209.59  7.8e-05   714  ! P58463 mus musculus (mouse).
SwissProt_40:FXP2_HUMAN +  196.00  206.84  0.0001   715  ! O15409 homo sapiens (human).
SwissProt_40:MCM1_YEAST +  192.00  210.37  0.0002   286  ! P11746 saccharomyces cerevis
SwissProt_40:KNQ1_YEAST +  189.50  199.24  0.0003   756  ! P53894 saccharomyces cerevis
SwissProt_40:SSN6_YEAST +  186.00  193.30  0.0005   966  ! P14922 saccharomyces cerevis
SwissProt_40:GLTA_WHEAT +  185.50  201.37  0.0005   356  ! P10385 triticum aestivum (wh
SwissProt_40:BRD4_HUMAN +  182.50  186.50  0.0008  1362  ! O60885 homo sapiens (human)
SwissProt_40:1NVO_PIG +  182.00  197.76  0.0007   347  ! P18175 sus scrofa (pig). invol
SwissProt_40:KAPC_DICDI +  181.50  191.82  0.0008   648  ! P34099 dictyostelium discoid
SwissProt_40:GALY_KLULA +  180.50  186.91  0.0010  1008  ! P32257 kluyveromyces lactis
SwissProt_40:ICP0_HSVBJ +  180.00  189.81  0.0010   676  ! P29128 bovine herpesvirus ty
SwissProt_40:GBF_DICDI +  180.00  189.41  0.0010   708  ! P36417 dictyostelium discoide
SwissProt_40:OVO_DROME +  178.00  184.00  0.0014  1028  ! P51521 drosophila melanogast
SwissProt_40:IXR1_YEAST +  177.50  188.15  0.0015   597  ! P33417 saccharomyces cerevis
SwissProt_40:KRUB_HUMAN -  176.50  196.76  0.0015   194  ! O75690 homo sapiens (human).
SwissProt_40:KCN3_HUMAN +  176.50  185.25  0.0017   736  ! Q9uqi6 homo sapiens (human).
SwissProt_40:YM38_YEAST +  176.50  184.99  0.0017   758  ! Q03825 saccharomyces cerevis
SwissProt_40:CIZ1_HUMAN +  176.00  182.98  0.0019   898  ! Q9ulv3 homo sapiens (human).
SwissProt_40:ICP0_HSVBK +  175.00  184.34  0.0021   676  ! P29836 bovine herpesvirus ty
SwissProt_40:GAL4_YEAST +  174.50  179.74  0.0024  1081  ! P19659 saccharomyces cerevi
SwissProt_40:FSH_DROME +  173.00  172.62  0.0031  2038  ! P13709 drosophila melanogast
SwissProt_40:ABF1_HUMAN +  172.50  166.92  0.0036  3703  ! Q15911 homo sapiens (human)
SwissProt_40:E78A_DROME +  172.00  178.94  0.0033   864  ! P45447 drosophila melanogast
SwissProt_40:KRUA_HUMAN -  171.50  192.48  0.0030   169  ! P26371 homo sapiens (human).
SwissProt_40:VIT_1CHUN +  171.50  171.94  0.0038  1823  ! Q91062 ichthyomyzon unicuspi
SwissProt_40:VIT1_CHICK +  171.50  171.53  0.0038  1912  ! P87498 gallus gallus (chick
SwissProt_40:SWI1_YEAST +  170.50  173.68  0.0043  1314  ! P09547 saccharomyces cerevi
SwissProt_40:MAM_DROME +  170.50  172.00  0.0043  1596  ! P21519 drosophila melanogast
SwissProt_40:TRX_DROVI +  170.50  164.45  0.0048  3828  ! Q24742 drosophila virilis (f
SwissProt_40:NAB2_YEAST +  169.00  179.96  0.0048   525  ! P32505 saccharomyces cerevis
SwissProt_40:ATX1_HUMAN +  169.00  176.15  0.0050   816  ! P54253 homo sapiens (human).
SwissProt_40:PRO_DROVI +  168.50  170.03  0.0057  1556  ! Q9u6a1 drosophila virilis (f
SwissProt_40:PHP_DROME +  168.50  169.85  0.0057  1589  ! P39769 drosophila melanogast
SwissProt_40:TCOF_HUMAN +  168.00  170.33  0.0061  1411  ! Q13428 homo sapiens (human)
SwissProt_40:SIMA_DROME +  168.00  169.76  0.0061  1507  ! Q24167 drosophila melanogas
SwissProt_40:YQC1_CAEEL +  167.50  179.31  0.0058   468  ! Q09260 caenorhabditis ele
SwissProt_40:RBF1_CANAL +  167.50  178.28  0.0059   527  ! Q00312 candida albicans (
SwissProt_40:NFT5_HUMAN +  167.50  169.07  0.0066  1531  ! O94916 homo sapiens (hum
SwissProt_40:ABF1_MOUSE +  167.00  160.85  0.0078  3726  ! Q61329 mus musculus (mou
SwissProt_40:YEAE_SCHPO +  166.00  176.76  0.0072   520  ! O14079 schizosaccharomyce

seq_name: SwissProt_40:YBM5_SCHPO

seq_documentation_block:
ID   YBM5_SCHPO     STANDARD;      PRT;   878 AA.
AC   Q10337;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical 98.4 kDa protein C582.05c in chromosome II.
GN   SPBC582.05C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   Wood V., Rajandream M.A., Barrell B.G., Davis P., Churcher C.M.;
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: SOME, TO YEAST YHR154W.
CC   -!- SIMILARITY: CONTAINS 5 BRCT DOMAINS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AL096788; CAB46668.1; -.
DR   InterPro; IPR001357; BRCT.
DR   Pfam; PF00533; BRCT; 6.
DR   SMART; SM00292; BRCT; 5.
DR   PROSITE; PS50172; BRCT; 5.
KW   Hypothetical protein; Repeat.
FT   DOMAIN       14    104       BRCT 1.
FT   DOMAIN      128    196       BRCT 2.
FT   DOMAIN      311    399       BRCT 3.
FT   DOMAIN      695    744       BRCT 4.
FT   DOMAIN      767    864       BRCT 5.
SQ   SEQUENCE   878 AA;  98435 MW;  AFCAFBEF800CBFC6 CRC64;


alignment_scores:
            Quality:  314.50              Length:   1194
              Ratio:  0.758                Gaps:     35
 Percent Similarity:  34.757   Percent Identity: 17.169

alignment_block:
 US-09-664-641-10 x YBM5_SCHPO   ..

 Align seg 1/1  to: YBM5_SCHPO  from: 1  to: 878

     144 TACGGATCTCTCGGCCCCTGCCCTTCAAATCACCTCTAGGCACAGGTTAT 193
         |||::::::|||   |||      :::|||:::|||      :::   ::
      25 TyrSerAlaLeuGlnPro...AsnGlnAsnGluLeuArgLysLysGluLe 40

     194 TCAGCTTCTCAAGGCTGGAAAAGCGAAGGAAGTTTCCTACAAT...GCAC 240
         :   :::   :::   |||||||||   :::      |||:::      |
      40 uPheIleLysAsnAspGlyLysAlaLeuSerPheProTyrAspTrpLysL 57

     241 TAGCCTCACACATAATCTCAGAGGATGGGGACAATCCAGAGGTGGGAGAA 290
         |||||:::|||:::|||   :::|||   ::::::|||:::|||   |||
      57 euAlaThrHisValIleCysAspAspPheSerSerProAsnValGlnGlu 73
```


291 GCTCGGGAAGTCTTTGACTTACCTGTTGTAAAGCCTTCTTGGGTGATTCT 340
74 GlySerLys...ArgSerLeuArgLeuAlaLysThrAsnTrpIleArgAs 89
341 GTCCGTTCAGTGTGGAACTCTTCTGCCAGTAAATGGTTTTTCTCCAGAAT 390
89 pCysValAspLysAsnThrLeuLeuAsnTyrSerPheTyrSerCysAsnP 106
391 CATGTCAGATTTTTTTTGGAATCACTGCCTGCCTTTCTCAGGGTGTTGAT 440
106 roTyrLeuLeuPheLysGlyIleCysAlaSerSerCysGln...IleAsp 121
441 ACAAGCTGGAGCTCTTTGTTGGAGTCTTCCAGAGCTCTCCCAAGGAGAGG 490
122 SerTyrGlnSerSerLeuIleAsp..... 129
491 TAGGGAAGGGAGCTTGTCCAGCAGAAGTTGGGAAGCACAGAGATCATCTG 540
129 129
541 CCTTCTTCTGACCCGGTATTGATGCAGGCTGAGGCCTCTGTTGTAATGTG 590
130 .....AspAla.LeuGluThr..... 134
591 CTGGGTGTCATCTGAAGACAGAAGTGCCCTGTGGGCTTTGGTTACGTTCT 640
135 .....P 135
641 ATGGGGGAGATTGCCAGCTAACCCTCAATAAGAAATGCACGCATTTGATT 690
135 heGlyGlyArgPheSerLysGlyLeuMetLysSerMetThrHisLeuPhe 151
691 GTTCCAGAGCCAAAGGGGGGGAAATACGAATGTGCTTTAAAGCGA..... 735
152 ThrTyrSerGlyMetGlyAlaLysCysLysLysValLeuAspLysProSe 168
736 .GCAAGTATTAAAATTGTGACTCCTGACTGGGTTCTGGATTGCGTATCAG 784
168 rLeuSerIleLysLeuIleHisProGlnTrpLeuLeuAspCysLeuGlnP 185
785 AGAAAACCAAAAAGGACGAAGCATTTTATCATCCTCGTCTGATTATTTAT 834
185 heGlyGlnLeuIleAspGln..... 191
835 GAAGAGGAAGAAGAGGAAGAGGAAGAGGAGGAGGAAGTAGAAAATGAGGA 884
191 191
885 ACAAGATTCTCAGAATGAGGGTAGTACAGATGAGAAGTCAAGCCCTGCCA 934
191 191
935 GCTCTCAAGAAGGGTCTCCTTCAGGTGACCAGCAGTTTTCACCTAAATCC 984
192 .....AspProTyrLeuPheProAsnPro 199
985 AACACTGAAAAATCTAAAGGGGAATTAATGTTTGATGATTCTTCAGATTC 1034
200 SerTyrLysLys.....AsnAspSe 206
1035 ATCACCGGAAAAACAGGAGAGAAATTTAAACTGGACCCCGGCCGAAGTCC 1084
206 rSerIleSerLysAlaGlu..... 212
1085 CACAGTTAGCTGCAGCAAAACGCAGGCTGCCTCAGGGAAAGGAGCCTGGG 1134
212 212
1135 TTGATTAACTTGTGTGCCAATGTCCCACCCGTCCCAGGTAACATTTTGCC 1184
213 .....ProThrSerLeu..... 216
1185 CCCTGAGGTCCGGGGTAATTTAATGGCTGCTGGACAAAACCTCCAAAGTT 1234


216 216
1235 CTGAAAGATCAGAAATGATAGCTACCTGGAGTCCAGCTGTACGGACACTG 1284
217 .....Phe 217
1285 AGGAATATTACTAATAATGCTGACATT.....CAGCAGATGAACCG 1325
218 ArgAsnValLeuHisGlyLysArgIleTyrPheSerAsnAspLeuAsnLe 234
1326 GCCATCAAATGTAGCACATATCTTACAGACTCTTTCAGCACCTACGAAAA 1375
234 uProThrAsnPheArgHisSerLeuGlnLys..... 244
1376 ATTTAGAACAGCAGGTGAATCACAGCCAGCAGGGACATACAAATGCCAAT 1425
244 244
1426 GCAGTGCTGTTTAGCCAAGTGAAAGTCACTCCAGAGACACACATGCTACA 1475
244 244
1476 GCAGCAGCAGCAGGCCCAGCAGCAGCAGCAGCAGCACCCGGTTTTACACC 1525
244 244
1526 TTCAGCCCCAGCAGATAATGCAGCTCCAGCAGCAGCAGCAGCAGATC 1575
244 244
1576 TCTCAGCAACCTTACCCCCAGCAGCCGCCGCATCCATTTTCACAGCAGCA 1625
244 244
1626 GCAGCAGCAACAGCAGCCACCACCATCGCCTCAGCAGCATCAGCTTTTTG 1675
244 244
1676 GACATGATCCAGCAGTGGAGATTCCAGAAGAAGGCTTCTTATTGGGATGT 1725
245 .....PheSerValGlyIle 249
1726 GTGTTTGCAATTGCGGATTATCCAGAGCAGATGTCTGATAAGCAACTGCT 1775
250 GlyAlaLysIleAla.....GluSerIleAsnAspCysAspIlePh 263
1776 GGCCACCTGGAAAAGGATAATCCAGGCACATGGCGGCACTGTTGACCCCA 1825
263 eIleGlyLeuLysArg.....AspThrIleGlu.... 272
1826 CCTTCACGAGTCGATGCACGCACCTTCTCTGTGAGAGTCAAGTCAGCAGC 1875
272 272
1876 GCGTATGCACAGGCAATAAGAGAAAGAAAGAGATGTGTTACTGCACACTG 1925
273 ...PheAsnLeuAlaSerAsnLysAsnThrThrIleGlyThrIleSerTr 288
1926 G...TTAAACACAGTCTTAAAAAAAAAAAAAAAGGTACCGCCGCACCGAG 1972
288 pLeuLeuAsnLeuPheValLeuGlySerTrpLysSerProLeuLeuAsnA 305
1973 CCCTTCACTTCCCAGTGGCCTTCCCACCAGGAGGAAAGCCATGTTCACAG 2022
305 laLeuHisTyrPro.....PheProSerValGlyPhe...LeuLysAsp 318
2023 CATATTATTTCTGTGACTGGATTTGTTGATAGTGACAGAGATGACCTAAA 2072
319 GlnMetValAlaValThrAsnTyrThrAspAlaAlaArgIleTyrLeuGl 335
2073 ATTAATGG TTATTTGGCAGGTGCCAAATATACGGGTTATCTATGCCGCA 2122

```
 335 uLysLeuLeuLeuAlaCysGlyAlaThrTyrThrLysAspLeuLysProT 352

2123 GCAACACAGTCCTCATCTGTAAAGAACCAACTGGTTTAAAGTATGAAAAA 2172
     ::|||||:::||||||:::   :::      |||   ||||||
 352 hrAsnThrLeuLeuIleAlaAlaSerSerTyrGlyGlnLysTyrGlyAla 368

2173 GCCAAAGAGTGGAGGATACCCTGTGTCAACGCCCAGTGGCTTGGCGACAT 2222
     ||||||   |||:::||||||   |||:::   :::||||||
 369 AlaLysValTrpAsnIleProThrValHisHisSerTrpLeu........ 382

2223 TCTTCTGGCAAACTTTGAGGCACTGAGGCACATTCAGTATAGTCGCTACA 2272
                :::::::::::::::::::::::::   :::   :::|
 383 .............TyrSerSerPheLysAsnLeuSerSerGlnAlaPheT 395

2273 CGGCATTC........................................... 2280
     ||   |||
 395 hrAspPheProValProLeuAspAspSerTyrMetAspPheIlePhePro 411

2280 ............................................... 2280

 412 CysProLeuAsnValGluLysGlySerPheGluAspThrLeuLysSerSe 428

2281 ..............................AGTCTGCAGGATCCA. 2295
                                     :::|||:::|||||
 428 rLeuThrLysGlyAsnSerGluValLeuLeuAspAspLeuSerAspProS 445

2295 ............................................... 2295

 445 erValSerSerIleLysGlyAsnLysThrAsnGluGluLeuGluLysGlu 461

2296 TTTGCCCCTACC..............CAGCATTTAGTTTTA........ 2322
     |||     |||            :::|||:::::|||
 462 PheLysSerThrSerAspAsnPheGlyLysHisIleIleLeuThrSerSe 478

2322 ............................................... 2322

 478 rPheSerAsnGlnSerAlaAspLysGlySerSerLeuAlaAlaGluAspA 495

2323 ..........................................AATCTTTTA 2331
                                               :::|||
 495 spArgAsnAspGluGlySerThrIleThrGlyValAsnArgGluLeuGln 511

2332 GATGCTTGGAGAGTT.................................. 2346
     |||      |||:::
 512 AspGluGlyArgLeuGluIleAspAlaLysSerSerLysThrAsnThrPr 528

2347 .......CCCTTAAAAGTGTCTGCAGAGTTGTTTCATGAGTATAAGA.... 2385
           ||||||   |||::::::            |||::::::
 528 oProSerProLeuLeuValGlyThrProSerLysGluSerLeuLysGluA 545

2386 ........................CTACCTCCCAAACTGAAACAGAAT 2409
                             |||      ||||||   :::|||
 545 laSerSerAspAspGluLeuProValLeuAlaThrLysLeuValAspAsn 561

2410 GAAGTAGCTAATGTCCAGCCTTCTTCCCAAAGAGCCAGAATT........ 2451
        :::   :::   :::|||   |||         ::::::
 562 ValIleLysGluLysSerProLeuSerLeuThrProLysValValValPr 578

2452 .......GAAGACGTACCACCTCCCACTAAAAAGCTAACTCCAGAATTG. 2493
           :::::::::            |||||||||      ||||||
 578 oSerHisLysGluThrTyrThrAspGluLysLysLeuIleAspGluLeuA 595

2493 ............................................... 2493

 595 spArgValAsnProLeuAsnSerSerGlnLeuLeuArgSerLysArgLys 611

2493 ............................................... 2493

 612 SerAlaAlaThrAlaLeuSerMetLeuGlnAsnValIleMetProAspVa 628

2493 ............................................... 2493

 628 lLeuAlaPheGluArgGluLysLysArgArgGlnThrHisArgSerValS 645

2493 ............................................... 2493

 645 erSerGlyGluValSerArgGluSerSerGluSerArgAsnThrAsnAla 661

2494 ............ACCCCTTTTGTGCTTTTCACTGGATTCGAGCCTGTCCA 2531
                        ::::::   |||||||||::::::
 662 LysAlaSerLysArgValTyrIleThrPheThrGlyTyrAspLysLysPr 678

2532 GGTTCAACAGTATATTAAGAAGCTCTACATTCTTGGTGGAGAGGTTGCGG 2581
           :::   :::|||||||||||   :::          ::::::::::
 678 oSerIleAspAsnLeuLysLysLeuAspMet.........SerIleThrS 692

2582 AGTCTGCACAGAAGTGCACACACCTCATTGCCAGCAAAGTGACTCGCACC 2631
     :::::   :::||||||||||||||||||||||   ::::::   ||||||
 692 erAsnProSerLysCysThrHisLeuIleAlaProArgIleLeuArgThr 708

2632 GTGAAGTTCCTGACGGCGATTTCTGTCGTGAAGCACATAGTGACGCCAGA 2681
        |||||||||   :::|||                  :::|||||||   ::
 709 SerLysPheLeuCysSerIleProTyrGlyProCysValValThrMetAs 725

2682 GTGGCTGGAAGAATGCTTCAGGTGTCAGAAGTTCATTGATGAGCAGAACT 2731
     ::|||:::::::::|||::::::   ::::::::::|||||||:::   |
 725 pTrpIleAsnSerCysLeuLysThrHisGluIleValAspGluGluProT 742

2732 ACATTCTCCGAGATGCTGAGGCAGAAGTACTTTTCTCTTTCAGCTTGGAA 2781
     ||:::|||:::|||   |||   |||:::   ::::::   :::||||||
 742 yrLeuLeuAsnAspProGluLysGluLeuGluLeuGlyCysThrLeuGlu 758

2782 GAATCCTTAAAACGGGCACACGTTTCT......CCACTCTTTAAGGCAAA 2825
     ::::::|||||||||||:::::::::         |||::::::
 759 SerAlaLeuLysArgAlaArgAlaGlnGlyProSerLeuLeuGluAspTy 775

2826 ATATTTTTACATCACA...CCTGGAATCTGCCCA...AGTCTTTCCACTA 2869
           |||:::|||         :::::::|||   ::::::   ::::
 775 rValValTyrLeuThrSerLysThrValAlaProGluAsnValProAlaV 792

2870 TGAAGGCAATCGTAGAGTGTGCAGGAGGAAAAGGTGTTATCCAAGCAGCCA 2919
     ::   :::||||||:::      ||||||         :::   :::
 792 alIleSerIleValLysSerAsnGlyGlyValCysSerThrLeuAsnVal 808

2920 TCTTTCCGGAAGCTCATGGAGCACAAGCAGAACTCGAGTTTGTCGGAAAT 2969
           ::::::|||   :::|||   ::::::::::::          ::
 809 TyrAsnLysArgLeuAlaArgHisLeuGluAspGlyAsn.........Va 822

2970 AATTTTAATATCCTGTGAAAATGACCTTCATTTATGCCGAGAATATTTT. 3018
     ::::||||||:::|||:::::|||   |||:::      :::::::::
 822 lValLeuIleThrCysAsnGluAspSerHisIleTrpThrAsnPheLeuA 839

3019 ..............GCCAGAGGCATAGATGTTCACAATGCAGAGTTCGTT 3054
                     :::   |||   :::::|||   :::::::::
 839 spAsnAlaSerGlnAsnLysThrIlePheLeuGlnAsnTyrAspTrpLeu 855

3055 CTGACTGGAGTGCTCACTCAAACGCTGGAC 3084
     :::      ||||||   |||   :::|||
 856 IleLysThrValLeuArgGlnGluIleAsp 865
```

seq_name: SwissProt_40:SNF5_YEAST

seq_documentation_block:

ID SNF5_YEAST STANDARD; PRT; 905 AA.
AC P18480;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcription regulatory protein SNF5 (SWI/SNF complex component SNF5)
DE (Transcription factor TYE4).
GN SNF5 OR TYE4 OR SWI10 OR YBR289W OR YBR2036.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=MCY;
RX   MEDLINE=91042489; PubMed=2233708;
RA   Laurent B.C., Treitel M.A., Carlson M.;
RT   "The SNF5 protein of Saccharomyces cerevisiae is a glutamine- and
RT   proline-rich transcriptional activator that affects expression of a
RT   broad spectrum of genes.";
RL   Mol. Cell. Biol. 10:5616-5625(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE=94378722; PubMed=8091861;
RA   Holmstroem K., Brandt T., Kallesoe T.;
RT   "The sequence of a 32,420 bp segment located on the right arm of
RT   chromosome II from Saccharomyces cerevisiae.";
RL   Yeast 10:S47-S62(1994).
CC   -!- FUNCTION: INVOLVED IN TRANSCRIPTIONAL ACTIVATION. THE SWI/SNF
CC       COMPLEX IS REQUIRED FOR THE INDUCED EXPRESSION OF A LARGE NUMBER
CC       OF GENES. THIS COMPLEX ALTERS CHROMATIN STRUCTURE TO FACILITATE
CC       BINDING OF GENE-SPECIFIC DEDICATED TRANSCRIPTION FACTORS.
CC   -!- SUBUNIT: COMPONENT OF THE SWI/SNF GLOBAL TRANSCRIPTION ACTIVATOR
CC       COMPLEX.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: BELONGS TO THE SNF5 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M36482; AAA35062.1; -.
DR   EMBL; X76053; CAA53652.1; -.
DR   EMBL; Z36158; CAA85254.1; -.
DR   PIR; S44551; RGBYS5.
DR   PIR; S39145; S39145.
DR   SGD; S0000493; SNF5.
KW   Transcription regulation; Activator; Nuclear protein.
FT   DOMAIN       31    270       GLN-RICH.
FT   DOMAIN       72    132       PRO-RICH.
FT   DOMAIN      272    324       PRO-RICH.
FT   DOMAIN      489    588       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN      714    882       PRO-RICH.
FT   DOMAIN      755    798       ARG/LYS RICH (BASIC).
FT   CONFLICT    564    564       E -> D (IN REF. 1).
SQ   SEQUENCE   905 AA;  102557 MW;  A287B4A648DD1A35 CRC64;


alignment_scores:
            Quality:  223.50            Length:    303
              Ratio:  1.380               Gaps:      9
 Percent Similarity:  53.465   Percent Identity: 27.723

alignment_block:
 US-09-664-641-10 x SNF5_YEAST   ..

 Align seq 1/1  to: SNF5_YEAST  from: 1  to: 905

     817 CCTCGTCTGATTATTTATGAACAGGAAGAAGAGGAAGAGGAAGAGGAGGA 866
         |||:::   :::         :::::::::::::::::      :::::
      42 ProGlnGlnLeuGlnMetIleGlnGlnArgHisGlnGlnLeuLeuArgSe 58

     867 GGAAGTAGAAAATGAGGAACAAGATTCTCAGAATGAGGGTAGTACAGATG 916
         :::::::::::::::::::|||::::::|||              :
      58 rArgLeuGlnGlnGlnGlnGlnGlnGlnGln.................G 69

     917 AGAAGTCAAGCCCTGCCAGCTCTCAAGAAGGGTCTCCTTCAGGTGACCAG 966
         ::   |||   |||   ::::::::::::::   |||         |||
      69 lnThrSerProProProGlnThrHisGlnSerProProProProProGln 85

     967 CAGTTTTCACCTAAATCCAACACTGAAAAATCTAAAGGGGAATTAATGTT 1016
         |||   :::|||   :::|||   :::
      86 GlnSerGlnProIleAlaAsnGlnSerAla..................... 95

    1017 TGATGATTCTTCAGATTCATCACCGGAAAAACAGGAGAGAAATTTAAACT 1066
              :::|||         |||           :::|||||:::
      96 .......ThrSerThrProProProProProAlaProHisAsnLeuHis. 109

    1067 GGACCCCGGCCGAAGTCCCACAGTTAGCTGCAGCAAAACGCAGGCTGCCT 1116
                      ||||||:::
     110 ................ProGlnIle....................... 112

    1117 CAGGGAAAGGAGCCTGGGTTGATTAACTTGTGTGCCAATGTCCCACCCGT 1166
           |||:::   |||                          :::   |||::
     113 ...GlyGlnValPro.....................LeuAlaProAl 120

    1167 CCCAGGTAACATTTTGCCCCCTGAGGTCCGGGGTAATTTAATGGCTGCTG 1216
         :|||   |||   ||||||||||:::::           :::|||:::
     120 aProIleAsn...LeuProProGlnIleAlaGlnLeuProLeuAlaThrG 136

    1217 GACAAAACCTCCAAAGTTCTGAAAGATCAGAAATGATAGCTACCTGGAGT 1266
           |||:::::::   ::::::   |||::::::   ||||||      :::
     136 lnGlnGlnValLeuAsnLysLeuArgGlnGlnAlaIleAlaLysAsnAsn 152

    1267 CCAGCTGTACGGACACTGAGGAATATTACTAATAATGCTGACATTCAGCA 1316
         |||   |||   :::     ::::::::::::           |||||
     153 ProGlnValValAsnAlaIleThrValAlaGln.............GlnGl 165

    1317 GATGAACCGGCCATCAAATGTAGCACATATCTTACAGACTCTTTCAGCAC 1366
         |:::::::|||      :::                |||||   ::::::
     165 nValGlnArgGlnIleGluGlnGlnLysGlyGlnGlnThrAlaGlnThrG 182

    1367 CTACGAAAAATTTAGAACAGCAGGTGAATCACAGCCAGCAGGGACATACA 1416
              ::::::   :::|||   :::   ::::::||||||   :::
     182 lnLeuGluGlnGlnArgGlnLeuLeuValGlnGlnGlnGlnGlnGlnGln 198

    1417 AATGCCAATGCAGTGCTGTTTAGCCAAGTGAAAGTGACTCCAGAGACACA 1466
               |||   :::      :::|||   :::         :::
     199 LeuArgAsnGlnIleGlnArgGlnGlnGlnGlnGlnPheArgHisHisVa 215

    1467 CATGCTACAGCAGCAGCAGCAGGCCCAGCAGCAGCAGCAGCAGCACCCGG 1516
         ::::::|||||||||||||||||   ||||||||||||||||||:::
     215 lGlnIleGlnGlnGlnGlnGlnLysGlnGlnGlnGlnGlnGln.... 230

    1517 TTTTACACCTTCAGCCCCAGCAGATAATGCAGCTCCAGCAGCAGCAGCAG 1566
              |||   |||   ||||||   :::|||   ||||||||||||||
     231 .....HisGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 245

    1567 CAGCAGATCTCTCAGCAACCTTACCCCCAGCAGCCGCCGCATCCATTTTC 1616
         ||||||   :::||||||         ||||||      :::       ::
     246 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGl 262

    1617 ACAGCAGCAGCAGCAGCAACAGCAGCCACCACCATCG.........CCTC 1657
         :|||||||||||||||||   |||   |||   |||          ||||
     262 nGlnGlnGlnGlnGlnGlnGlyGlnIleProGlnSerGlnGlnValProG 279

    1658 AGCAGCATCAGCTTTTTGGACATGATCCAGCAGTGGAGATTCCAGAAGAA 1707
         ||   :::::::::   |||:::   |||:::         |||
     279 lnValArgSerMetSerGlyGlnProProThrAsnValGlnProThrIle 295

    1708 GGCTTCTTA 1716
         |||   |||
     296 GlyGlnLeu 298

seq_name: SwissProt_40:YKF4_YEAST

seq_documentation_block:
ID   YKF4_YEAST     STANDARD;      PRT;   738 AA.
AC   P35732;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   Hypothetical 84.0 kDa protein in NUP120-CSE4 intergenic region.
GN   YKL054C OR YKL308.
```

```
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE=94378723; PubMed=8091862;
RA   Rasmussen S.W.;
RT   "Sequence of a 28.6 kb region of yeast chromosome XI includes the
RT   FBA1 and TOA2 genes, an open reading frame (ORF) similar to a
RT   translationally controlled tumour protein, one ORF containing motifs
RT   also found in plant storage proteins and 13 ORFs with weak or no
RT   homology to known proteins.";
RL   Yeast 10:S63-S68(1994).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X75781; CAA53418.1; -.
DR   EMBL; Z28054; CAA81890.1; -.
DR   PIR; S37876; S37876.
DR   PIR; S39184; S39184.
DR   SGD; S0001537; YKL054C.
KW   Hypothetical protein.
FT   DOMAIN      382    718       GLN-RICH.
FT   DOMAIN      574    580       POLY-ALA.
FT   DOMAIN      708    716       POLY-ALA.
SQ   SEQUENCE   738 AA;  83973 MW;  05734E2D0B7389AC CRC64;


alignment_scores:
            Quality:  213.00            Length:    324
              Ratio:  1.275               Gaps:     11
 Percent Similarity:  51.543   Percent Identity: 26.543

alignment_block:
 US-09-664-641-10 x YKF4_YEAST   ..

 Align seq 1/1  to: YKF4_YEAST  from: 1  to: 738

     835 GAAGAGGAAGAAGAGGAAGAGGAAGAGGAGGAGGAAGTAGAAAATGAGGA 884
         |||:::|||:::||||||:::::::::::::|||      :::|||:::||
     199 GluLysGluGlnGluGluGlnGlnGlnGlnGluGlyHisAsnAsnLysGl 215

     885 ACAAGATTCTCAGAATGAG.............................. 903
         |:::   :::|||   |||
     215 uGluHisLysGlnIleGluGlnProSerLeuSerSerLysLysThrThrS 232

     904 ..........GGTAGTACAGATGAGAAGTCAAGCCCTGCCAGCTCTCAA 942
                 :::|||      :::|||   |||   |||:::
     232 erArgThrSerAlaSerGlnProLysLysMetSerTrpAlaAlaIleAla 248

     943 GAAGGGTCTCCTTCAGGTGACCAGCAGTTT...TCACCTAAATCCAACAC 989
             :::|||::::::   ::::::      ||||||   :::|||::
     249 ThrProLysProLysAlaValLysLysThrGluSerProLeuGluAsnVa 265

     990 TGAAAAATCTAAAGGGGAATTAATGTTTGATGATTCTTCAGATTCATCAC 1039
         :   :::   |||   |||:::           ::::::|||::::::
     265 lAlaGluLeuLysLysGluIleSerAspIleLysLysAspAspGlnLysS 282

    1040 CCGAAAAACAGGAGAGAAATTTAAACTGCACCCCGGCCGAAGTCCCACAG 1089
           |||   :::|||::::::::::|||                   |||
     282 erGluAlaSerGluGluLysValAsn.................GluGln 292

    1090 TTAGCTGCAGCAAAACGCAGGCTGCCTCAGGGAAAGGAGCCTGGGTTGAT 1139
         ::::::|||::::::::::     :::      |||||||:::
     293 GluThrSerAlaGlnGluGlnGluGluGluThrAlaGluProSerGluGl 309

    1140 TAACTTGTGTGCCAATGTCCCACCCGTCCCAGGTAACATTTTGCCCCCTG 1189
          |||         :::||||||   |||   |||:::           |
     309 uAsnGlu...AspArgValProGluValAspGlyGlu.............G 321

    1190 AGGTCCGGGGTAATTTAATGGCTGCTGGACAAAACCTCCAAAGTTCTGAA 1239
         |||||:::                                 ::::::::|||
     321 luValGln.............................GluGluAlaGlu 327

    1240 AGATCAGAAATGATAGCTACCTGGAGTCCAGCTGTACGGACACTGAGGAA 1289
         ::::::|||::::::         :::  :::::::::   |||:::::
     328 LysLysGluGlnValLysGluGluGluGlnThrAlaGluGluLeuGluGl 344

    1290 TATTACTAATAAT......GCTGACATTCAGCAGATGAACCGGCCATCAA 1333
         :      :::|||      |||      ::::::::::::      ::::
     344 nGluGlnAspAsnValAlaAlaProGluGluGluValThrValValGluG 361

    1334 ATGTAGCACATATCTTACAGACTCTTTCAGCACCTACGAAAAATTTAGAA 1383
         ::   ::::::|||      ::::::|||   |||   ::::::
     361 luLysValGluIleSerAlaValIleSerGluProProGluAspGlnAla 377

    1384 CAGCAGGTGAATCACAGCCAGCAGGGACATACAAATGCCAATGCAGTGCT 1433
         :::   |||   :::   ||||||
     378 AsnThrValProGlnProGlnGln..................................... 385

    1434 GTTTAGCCAAGTGAAAGTGACTCCAGAGACACACATGCTACAGCAGCAGC 1483
                |||   :::         :::   ::::::   ||||||   |
     386 .......GlnSerGlnGlnProGlnGlnProGlnGlnProGlnGlnProG 400

    1484 AGCAGGCCCAGCAG...CAGCAGCAGCAGCACCCGGTTTTACACCTTCAG 1530
         |||||   |||||   ||||||||||||:::|||            |||
     400 lnGlnProGlnGlnProGlnGlnGlnGlnGlnProGlnGlnProGlnGln 416

    1531 CCCCAGCAGATAATGCAGCTCCAGCAGCAGCAGCAGCAG.......... 1569
         |||||||||   :::|||   ||||||||||||||||||
     417 ProGlnGlnGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnProValGl 433

    1570 .........................CAGATCTCTCAGCAACCTTACCCCC 1594
                                  |||:::||||||:::   |||   |
     433 nAlaGlnAlaGlnAlaGlnGluGluGlnLeuSerGlnAsnTyrTyrThrG 450

    1595 AGCAGCCGCCGCATCCATTTTCACAGCAG..................... 1623
         |||||      :::   ::::::||||||
     450 lnGlnGlnGlnGlnGlnTyrAlaGlnGlnGlnHisGlnLeuGlnGlnGln 466

    1624 .........................CAGCAGCAGCAACAGCAGCCACC 1646
                                  ||||||||||||||:::|||
     467 TyrLeuSerGlnGlnGlnGlnTyrAlaGlnGlnGlnGlnGlnHisProGl 483

    1647 ACCATCGCCTCAGCAGCATCAG 1668
          |||:::   ||||||   |||
     483 nProGlnSerGlnGlnProGln 490

seq_name: SwissProt_40:YO83_CAEEL

seq_documentation_block:
ID   YO83_CAEEL     STANDARD;      PRT;  1000 AA.
AC   P34619;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Hypothetical 113.4 kDa protein ZK1236.3 in chromosome III.
GN   ZK1236.3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
```

```
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA   Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA   Johnston L., Jones M., Kershaw J., Kirsten J., Laisster N.,
RA   Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Sims M., Smaldon N., Smith A., Smith M., Sonnhammer E., Staden R.,
RA   Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA   Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA   Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L13200; AAA28193.1; -.
DR   PIR; S44898; S44898.
DR   WormPep; ZK1236.3; CE00532.
KW   Hypothetical protein.
SQ   SEQUENCE   1000 AA;  113419 MW;  D230C9A6F84928E2 CRC64;


alignment_scores:
            Quality:  204.50              Length:   1006
              Ratio:  0.492                Gaps:     47
 Percent Similarity:  41.352   Percent Identity: 19.682

alignment_block:
 US-09-664-641-10 x YO83_CAEEL   ..

 Align seg 1/1  to: YO83_CAEEL  from: 1  to: 1000

     139 AGGGCTACGGATCTCTCGGCCCCTGCCCTTCAAATCACCTCTAGGCACAG 188
         |||      |||  |||  |||  |||:::     |||:::
     104 ArgMetLysAspThrSerLeuProProLeuGluAsnProSerAsnGlyGl 120

     189 GTTATTCAGCTTCTCAAGGCTGGAAAAGCGAAGCAAGTTTCCTACAATGC 238
             :::   |||:::   :::               ||||||::::::
     120 yThrTyrLeuPheAsnLeuIle.........ThrValPheProSerIleA 134

     239 ACTAGCCTCACACATAATCTCAGAGGATGGGGACAATCCAGAGGTGGGAG 288
                 :::   |||   |||      |||           :::
     134 laLeuSerPheAsnTyrSerTrpArgGlyAspThrGlyProSerIleSer 150

     289 AAGCTCGGGAAGTCTTTGACTTACCTGTTGTAAAGCCTTCTTGGGTGATT 338
         ::::::   ||||||:::   |||:::   |||||||:::
     151 IlePheSerPheSerLeuSerValLeuPhePheSerLeuPheProGlnHi 167

     339 CTGTCCGTTCAGTGTGGAACTCTTCTGCCAGTAAATGGTTTTTCTCCAGA 388
               :::   ::::::   :::|||
     167 sLysAsnIleCysAlaArgAlaTrpCys..................... 176

     389 ATCATGTCAGATTTTTTTTGGAATCACTGCCTGCCTTTCTCAGGGTGTTG 438
             |||   |||   :::||||||        ||||||   :::
     177 ........ArgProPheArgArgSerLeuSerPheSerLeuArgPheIle 190

     439 ATACAAGCTGGAGCTCTTTGTTGGAGTCTTCCAGAGCTCTCCCAAGGAGA 488
         ::::::                  |||   |||      |||:::   ||
     191 MetSer.................SerGluProAlaSerSerSerThrGl 201

     489 G...GTAGGGAAGGGAGCTTGTCCAGCAGAAGTTGGGAAGCACAGAGATC 535
         |   |||   :::         |||   ::::::   ||||||:::   :
     201 uLysValProGluGluProHisProHisSerIle...LysHisLysPheG 217

     536 ATCTGCCTTCTTCTGACCCGGTATTGATGCAGGCTGAGGCCTCTGTTGTA 585
         ::   |||:::         |||:::   :::|||   ::::::   |||
     217 lnGlyProGlnPhe......ValIleProArgAlaLeuSerAspTyrVal 231

     586 ATGTGCTGGGTGTCATCTGAAGACAGAAGTGCCCTGTGGGCTTTGGTTAC 635
         :::         ::::::      :::|||         ||||||   ::
     232 LeuAsnValAsnAsnGlnThrProGluSerTyrGluLysAlaLeuAsnAl 248

     636 GTTCTATGGGGGAGATTGCCAGCTAACCCTCAATAAGAAATGCACGCATT 685
         :   ||||||   |||      :::||||||         |||   |||:
     248 aLysTyrGlyArgAspAspTyrIleThrLeu.........CysLeuHisV 262

     686 TGATTGTT.................CCAGAGCCAAAGGGGGGGAAATAC 717
         ::                        |||         |||   :::|||
     262 alThrSerLeuCysThrGluTyrGlyProLeuAspAspGlyProGluTyr 278

     718 GAA......TGTGCTTTAAAGCGAGCAAGTATTAAAATTGTG........ 753
                 |||      :::           |||:::::::
     279 ValLeuLeuCysAspSerArgAlaArgValAspLysLeuIleGluAspLe 295

     753 ..................................................... 753

     295 uValLysLeuLeuGluIleAspThrAspTyrValGlnLeuGluLeuHisG 312

     754 ......................ACTCCTGACTGGGTTCTGGATTGCGTA 780
                                  ||||||  ||||||      :::
     312 lyGlyLysArgLeuHisLeuGlnLysProAspAlaValLeuArgAspIle 328

     781 TCAGAGAAAACCAAA.........AAGGACGAAGCATTTTATCATCCTCG 821
         :::   |||   :::         :::|||:::   |||   :::   ::
     329 AlaTyrLysGlnSerAsnGluGlySerAspLysPhePheLeuGluMetLy 345

     822 TCTGATT.......................ATTTATGAAGAGGAAGAAG 847
         :|||:::                      |||   ::::::||||||
     345 sLeuValProSerGluSerMetLysAlaLysIleMetLysGlnGluGluG 362

     848 AGGAAGAG.......................GAAGAGGAGGAGGAAGTA 873
         ||||||||                      ::::::   :::|||
     362 luGluGluLysAlaArgLysHisGlyGlnTyrGlnGlnTyrGlnGluTyr 378

     874 GAAAATGAGGAACAA.........GATTCTCAGAATGAGGGTAGTACAGA 914
         :::::::::::::|||         |||:::|||::::::::::|||:::
     379 HisGlnGlnHisGlnAlaMetAsnAspGlyGlnSerSerSerSerValPr 395

     915 TGAGAAGTCAAGCCCTGCCAGCTCTCAAGAAGGGTCT.............. 951
           :::   |||||||||:::   |||:::|||::::::
     395 oSerThrSerSerProSerCysSerSerGluAlaAsnArgLysGluMetG 412

     952 ..............CCTTCAGGTGACCAGCAGTTT.............. 972
                       |||:::|||   :::::::::
     412 luThrValArgGluProAlaGlyProSerGluLeuMetArgAlaIleAsn 428

     973 .....................................TCACCTAAATC 983
                                              ::::||   ::
     429 AlaProValAlaProAlaProValValIleLysIleGluThrProValAl 445

     984 CAACACTGAAAAATCTAAAGGGGAATTAATGTTTGATGATTCTTCAGATT 1033
         :     |||:::::::::      ||||||   ||||||:::      |
     445 aLeuProGluGluAspGlu...ThrLeuMetAspAspAspGluMetProS 461

    1034 CA...................................................... 1035
         ||
     461 erLeuThrValGluAlaProSerGluGluAlaSerPheGluAlaGluGln 477

    1036 ...TCACCGGAAAAACAGGAGAGAAATTTAAACTGG......ACCCCGGC 1076
            ||||||:::      :::   :::::::::
     478 ProSerProGlnValProGlnAlaSerIleGluGlyProSerGlnGlnGl 494

    1077 CGAAGTCCCACAGTTAGCTGCAGCAAAACGCAGGCTGCCTCAGGGAAAGG 1126
          :::::::|||      :::      ||||||::::::   :::|||::::
     494 nGlnIleProGlyThrSerGlnGlnLysArgGlnValAlaArgGlySerA 511

    1127 AGCCTGGGTTGATTAACTTGTGTGCCAATGTCCCACCCGTCCCAGGTAAC 1176
         ::   ::::::|||:::         ::::::||||||      ||||||
     511 rgThrAsnMetIleSerTyr...HisAspLeuProProGlyThrGlyAsn 526
```

```
1177 ATTTTGCCCCCTGAGGTCCGGGGTAATTTAATGGCTGCTGGACAAAACCT 1226
          ||||||                    ||||||:::   |||
 527 ...AlaProPro.................MetAlaCysProGlnValTh 536

1227 CCAAAGTTCTGAAAGATCAGAAATGATAGCTACCTGGAGTCCAGCTGTAC 1276
       :::   |||::::::          :::   ::::::      :::|
 536 rLeuLysLeuGluLysAsn.........ValProPheGluAlaLysIleA 550

1277 GGACACTGAGGAATATTACTAATAAT......GCTGACATTCAGCAGATG 1320
     ||::::::   :::   |||:::::::      :::::::::|||:::|||
 550 rgAlaValAlaGlyTyrThrArgLysProIleSerGluValGlnLysMet 566

1321 AACCGGCCATCAAATGTAGCACATATCTTACAGACTCTTTCAGCACCTAC 1370
        |||||||||::::::      |||:::::::::::::
 567 ...ArgProSerAspLeuGluSerIlePheHisSerIle..........  578

1371 GAAAAATTTAGAACAGCAGGTGAATCACAGCCAGCAGGGACATACAAATG 1420

 578 ................................................... 578

1421 CCAATGCAGTGCTGTTTAGCCAAGTGAAAGTGACTCCAGAGACACACATG 1470
               :::   |||      ::::::      ::::::::::::
 579 ........CysIleAlaSerValGlnArgIleLysArgArgAsnGluLeu 592

1471 CTACAGCAGCAGCAGCAGGCCCAGCAGCAGCAGCAGCAGCACCCGGTTTT 1520
     :::||||||   |||:::   :::   |||:::   :::   |||:::::
 593 ValGlnGlnLeuGlnGluIleAsnAlaGlnSe*CysLysSerProThrMe 609

1521 ACACCTTCAGCCC.....................CAGCAGATAATG. 1545
     :   ::::::                        |||:::::::::
 609 tThrMetAsnLysLysPheThrLeuAlaLysAlaTyrGlnArgValGlnA 626

1546 ..CAGCTCCAGCAGCAGCAGCAGCAGCAGATCCCTCAGCAACCTTACCCC 1593
       ::::::::::::   ::::::::::||||||      |||   |||
 626 snGluIleGluLysIleAspArgGluGlnIleLeuProGlnGlnTyrMet 642

1594 CAGCAGCCGCCGCATCCATTTTCACAGCAGCAGCAGCAGCAACAGCAGCC 1643
     ::::::||||||   |||            |||   |||||||:::   ||
 643 AsnMetProProMetPro.........ProGlnGlyGlnGlnArgLeuPr 656

1644 ACCACCATCG.....................CCTCAGCAGCATCAGCTTT 1672
     |||||||:::                    ||||||||||::::::
 656 oProProAlaTyrProProGlyIleLeuProProGlnGlnAsnArgGlnG 673

1673 TTGGACATGATCCAGCAGTGGAGATTCCAGAAGAAGGCTTCTTATTGGGA 1722
       |||      |||      :::          :::   |||::::::|||
 673 lnGlyValProProGlnPheGlnArgSerProGln...PheMetIleGly 688

1723 TGTGTGTTTGCAATTGCGGATTATCCAGAGCAGATGTCTGATAAGCAACT 1772
                                          |||   |||
 689 ................................ProAspGlyGlnAr 693

1773 GCTGGCCACCTGGAAAAGGATAATCCAGGCACATGGCGGCACTGTTGACC 1822
                                 ||||||              |
 693 q......................TyrAlaHisProTyrMetGlnLeuP 702

1823 CCACCTTCACGAGTCGATGCACGCACCTTCTCTGTGACAGTCAAGTCAGC 1872
     ||:::   :::::::|||   ::::::::::|||      |||:::|||:::
 702 roAsnSerAsnGlnArg...AlaArgIleLeuAsnThrSerSerValGln 717

1873 AGCGCGTATGCACAGGCAATAAGAGAAAGAAAGAGATGTGTTACTGCACA 1922
               ::::::   :::|||:::|||         :::   |||
 718 Pro......SerGluGluValArgAsnArgLeuValLysIleGluAlaMe 732

1923 CTGGTTAAACACAGTCTTA............AAAAAAAAAAAAATGGTAC 1960
       ::::||   :::                :::               |
 732 tAlaMetAsnMetLAlaGlnLeuAsnProProArgProProProProGlnP 749

1961 CGCCGCACCGAGCCCTTCACTTCCCAGTGGCCTTC......CCACCAGGA 2004
     ||||||||||||||||||:::      :::   |||      |||   :::
 749 roProHisArgAlaLeuGlnGlyGluLeuGlnPheLeuArgProGlyAla 765

2005 GGAAAGCCATGTTCACAGCATATTATTTCTGTGACTGGATTTGTTGATAG 2054
```

```
          ||||||:::
 766 ProAspProCysAsn..................................  770

2055 TGACAGAGATGACCTAAAATTAATGGCTTATTTGGCAGGTGCCAAATATA 2104
                                     :::         ::::||
 771 ............................PheArgProAspSerLysGlnT 778

2105 CGGGTTATCTATGCCGCAGCAACACAGTCCTCATCTGTAAAGAACCAACT 2154
     ||               :::|||||   :::         :::|||:::
 778 hr............TyrAsnAsnThrTyrValThrValAlaSerProAla 790

2155 GGTTTA............................AAGTATGAAAAAGC 2174
       |||                                |||||||||::
 791 ThrLeuThrAsnSerIleIleProTrpHisPheProProTyrGluLysSe 807

2175 C.................................AAAGAGTGGAGGA 2188
     :                                  :::|||:::|||:
 807 rGlyArgLeuAsnValSerAsnThrIleLysAlaIleAsnGluTyrArgL 824

2189 TACCCTGTGTCAACGCCCAG.............................. 2208
     ::   |||   :::   |||
 824 euLeuCysAsnSerArgGlnAlaAspProAlaSerPheLeuGluPheTyr 840

2209 TGGCTTGGCGACATTCTTCTGGGAAACTTTGAGGCACTGAGGCAGATTCA 2258
     :::||||||||   :::          :::   |||        ::
 841 PheLeuGlyAspProMetProHisPheAsnLysIleLeuSerIleAlaAs 857

2259 GTATAGTCGCTACACGGCATTCAGTCTGCAGGATCCA.............. 2295
     :|||:::   |||   :::          |||
 857 pTyrAsnMetTyrLeuSerArgArgArgCysAspGluAlaAspValLysI 874

2296 ........TTTGCCCCTACCCAGCATTTAGTTTTAAATCTTTTAGAT... 2334
             ::::::   ::::::::::|||   |||   |||||||:::
 874 leHisArgMetSerHisSerAspGlnLeuGlnLeuTyrLeuLeuGluLeu 890

2334 ................................................... 2334

 891 GlnSerAspGluSerAsnValGluLysTrpLysThrPheTyrArgIleMe 907

2335 .GCTTGGAGAGTTCCCTTA............................... 2352
        |||   :::|||||||
 907 tGlnTrpAspLeuProLeuAsnAsnGluPheProArgIleLeuLeuProS 924

2353 .....................................AAAGTGTCTGCA 2364
                                         |||   |||
 924 erSerLeuAspIleGlyArgProValValAspArgLysLysLysSerIle 940

2365 GAGTTGTTGATGAGT.................ATAAGACTACCTCCCAA 2396
     :::   :::|||:::                :::   ||||||::
 941 AspGlnValMetAsnHisIleHisArgMetHisSerGlnArgProProSe 957

2397 ACTGAAACAGAATGAA................................ 2412
     ::::   :::::::::
 957 rMetGlyAsnSerSerThrSerSerGluAlaSerSerThrSerProThrA 974

2413 ........GTAGCTAATGTCCAGCCTTCTTCCCAAAGAGCCAGAATTGAA 2454
             :::|||:::   :::|||:::|||:::|||        :::
 974 snAlaAlaThrAlaThrSerSerProAlaSerAsnArgProThrThrSer 990

2455 GACGTACCACCTCCCACT 2472
       :::   |||||||||
 991 ThrAlaGlnProProThr 996
```

seq_name: SwissProt_40:Y192_HUMAN

seq_documentation_block:

```
ID   Y192_HUMAN     STANDARD;      PRT;  2124 AA.
AC   Q93074;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetica  protein KIAA0192 (Fragment).
GN   KIAA0192.
```

```
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow;
RX   MEDLINE=96281124; PubMed=8724849;
RA   Nagase T., Seki N., Ishikawa K.-I., Tanaka A., Nomura N.;
RT   "Prediction of the coding sequences of unidentified human genes. V.
RT   The coding sequences of 40 new genes (KIAA0161 KIAA0200) deduced by
RT   analysis of cDNA clones from human cell line KG-1.";
RL   DNA Res. 3:17-24(1996).
CC   -!- TISSUE SPECIFICITY: UBIQUITOUS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D83783; BAA12112.1; -.
DR   MIM; 300188; -.
KW   Hypothetical protein.
FT   NON_TER       1      1
FT   DOMAIN      599    602       POLY-SER.
FT   DOMAIN     1201   1207       POLY-GLY.
FT   DOMAIN     1998   2124       GLN-RICH.
FT   DOMAIN     1998   2023       POLY-GLN.
FT   DOMAIN     2028   2033       POLY-GLN.
FT   DOMAIN     2037   2070       POLY-GLN.
FT   DOMAIN     2090   2097       POLY-GLN.
SQ   SEQUENCE   2124 AA;  237207 MW;  255FB9419EC39F42 CRC64;


alignment_scores:
            Quality:  200.00              Length:    500
              Ratio:  0.952                Gaps:     22
 Percent Similarity:  42.000   Percent Identity: 23.200

alignment_block:
 US-09-664-641-10 x Y192_HUMAN   ..

 Align seg 1/1  to: Y192_HUMAN  from: 1  to: 2124

     658 CTAACCCTCAATAAGAAATGCACGCATTTGATTGTTCCAGAGCCA..... 702
         |||   |||   |||:::          :::|||:::   ||||||
    1586 LeuProLeuProLysGlnThrArgAspValIleThrCysGluProGlnGl 1602

     703 ..............AAGGGGGGGAAA.........TACGAATGTGCTT 727
                       ||||||:::|||         ::::::      :
    1602 ySerLeuIleAspThrLysGlyAsnLysIleAlaGlyPheAspSerIleP 1619

     728 TAAAGCGAGCAAGTATTAAAATTGTGACTCCTGAC............TGG 765
         ::|||:::   ::::::::::::   |||   :::            |||
    1619 heLysLysGluGlyLeuGlnValSerThrLysGlnLysIleSerProTrp 1635

     766 GTTCTG.......................................... 771
             |||
    1636 AspLeuPheGluGlyLeuLysProSerAlaProLeuSerTrpGlyTrpPh 1652

     772 .......GATTGCGTATCAGAGAAAACCAAAAAGGACGAA.......... 804
                  |||:::::::::::::   :::   |||
    1652 eGlyThrValArgValAspArgArgValAlaArgGlyGluGluGlnGlnA 1669

     805 .....GCATTTTATCAT............CCTCGTCTGATTATTTATGAA 837
              :::||||||             ||||||          |||
    1669 rgLeuLeuLeuTyrHisThrHisLeuArgProArgProArgAlaTyrTyr 1685

     838 GAGGAA.................GAAGAGGAAGAGGAAGAGGAGGAGGA 869
             |||                 |||:::||||||

    1686 LeuGluProLeuProLeuProProGluAspGluGluProProAlaProTh 1702

     870 AGTAGAAAATGAGGAACAAGATTCTCAGAATGAGGGTAGTACAGATGACA 919
          :::   :::   |||:::   :::   :::       :::|||||||:::
    1702 rLeuLeuGluProGluLysLysAlaProGluProProLysThrAspLysP 1719

     920 AG...TCAAGCCCTGCCAGCTCTCAAGAA.........GGGTCTCCTTCA 957
             ::::::|||   |||:::::::|||           |||   :::
    1719 roGlyAlaAlaProProSerThrGluGluArgLysLysLysSerThrLys 1735

     958 GGTGACCAGCAGTTTTCACCTAAATCCAACACTGAAAAA........... 996
         |||   ::::::   :::|||   ::::::||||||
    1736 GlyLysLysArgSerGlnProAlaThrLysThrGluAspTyrGlyMetGl 1752

     997 ..............TCTAAAGGGGAATTAATGTTTGATGATTCTTCAG 1030
                              |||       :::       |||
    1752 yProGlyArgSerGlyProTyrGlyValThrValProProAspLeuLeuH 1769

    1031 ATTCATCACCGGAAAAACAGGAGAGAAATTTAAACTGGACCCCGGCCGAA 1080
                         :::      :::|||||||:::      ::::::
    1769 isHisProAsnProGlySerIleThrHisLeuAsnTyrArgGlnGlySer 1785

    1081 GTC.....................CCACAGTTAGCTGCAGCAAAACGC.. 1107
         :::                     |||      |||::::::   |||
    1786 IleGlyLeuTyrThrGlnAsnGlnProLeuProAlaGlyGlyProArgVa 1802

    1108 ..................AGGCTGCCTCAGGGAAAGGAGCCTGGGTTGA 1138
                           |||||||||:::   |||   |||
    1802 lAspProTyrArgProValArgLeuProMetGlnLysLeuPro....... 1816

    1139 TTAACTTGTGTGCCAATGTCCCACCCGTCCCAGGTAACATTTTGCCCCCT 1188
                      :::   |||      ||||||   :::|||||||
    1817 ........ .....ThrArgProThrTyrProGly...ValLeuProThr 1827

    1189 GAGGTCCGGGGTAATTTA............................... 1206
            :::   |||   :::
    1828 ThrMetThrGlyValMetGlyLeuGluProSerSerTyrLysThrSerVa 1844

    1207 .....................ATGGCTGCTGGACAAAACCTCCAA.... 1230
                                :::      |||||||:::|||:::
    1844 lTyrArgGlnGlnGlnProAlaValProGlnGlyGlnArgLeuArgGlnG 1861

    1231 ............................AGTTCTGAAAGATCAGAAATG 1251
                                     :::|||:::      :::|||
    1861 lnLeuGlnGlnSerGlnGlyMetLeuGlyGlnSerSerValHisGlnMet 1877

    1252 ATAGCTACCTGGAGTCCAGCTGTACGGACACTGAGGAATATTACT..... 1296
             :::   |||   ::::::::::|||   ::::::   |||
    1878 ThrProSerSerSerTyrGlyLeuGlnThrSerGlnGlyTyrThrProTy 1894

    1297 ....AATAATGCTGACATTCAGCAG.....................ATGA 1321
         :::::::::   :::||||||                        |||
    1894 rValSerHisValGlyLeuGlnGlnHisThrGlyProAlaGlyThrMetV 1911

    1322 ACCGGCCATCAAATGTAGCACATATCTTACAG................. 1353
             ||||||      ::::::      |||
    1911 alProProSerTyrSerSerGlnProTyrGlnSerThrHisProSerThr 1927

    1354 .... .ACTCTTTCAGCACCTACGAAAAATTTAGAACAGCAGGTGAAT.. 1395
             ||||||      ||||||:::::::|||:::|||:::   :::
    1928 AsnProThrLeuValAspProThrArgHisLeuGlnGlnArgProSerGl 1944

    1396 .......CACAGCCAG................................ 1404
                |||:::|||
    1944 yTyrValHisGlnGlnAlaProThrTyrGlyHisGlyLeuThrSerThrG 1961

    1404 ............................................... 1404

    1961 lnArgPheSerHisGlnThrLeuGlnGlnThrProMetIleSerThrMet 1977

    1405 .......................CAGGGACATACAAATGCCAATGCAGT 1430
                                |||:::   ::::::::::::|||::
    1978 ThrProMetSerAlaGlnGlyValGlnAlaGlyValArgSerThrAlaIl 1994
```

```
1431 GCTGTTTAGCCAAGTGAAAGTGACTCCAGAGACACACATGCTACAGCAGC 1480
1994 eLeuProGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnG 2011

1481 AGCAGCAGGCCCAGCAGCAGCAGCAGCAGCACCCGGTTTTACACCTTCAG 1530
2011 lnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnTyrHisIleArg 2027

1531 CCCCAGCAGATAATGCAGCTC.....................CAGCAGCA 1559
2028 GlnGlnGlnGlnGlnGlnIleLeuArgGlnGlnGlnGlnGlnGlnGlnGl 2044

1560 GCAGCAGCAGCAGATCTCTCAGCAACCTTACCCCCAGCAGCCGCCGCATC 1609
2044 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnG 2061

1610 CATTTTCACAGCAGCAGCAGCAGCAACAGCAGCCACCACCATCGCCTCAG 1659
2061 lnGlnHisGlnGlnGlnGlnGlnGlnGlnAlaAlaProProGlnProGln 2077
```

seq_name: SwissProt_40:FXP2_MOUSE

seq_documentation_block:

ID FXP2_MOUSE STANDARD; PRT; 714 AA.
AC P58463;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Forkhead box protein P2.
GN FOXP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Lung;
RX MEDLINE=21347947; PubMed=11358962;
RA Shu W., Yang H., Zhang L., Lu M.M., Morrisey E.E.;
RT "Characterization of a new subfamily of winged-helix/forkhead (Fox)
RT genes that are expressed in the lung and act as transcriptional
RT repressors.";
RL J. Biol. Chem. 276:27488-27497(2001).
CC -!- FUNCTION: Transcriptional repressor that play an important role in
CC the specification and differentiation of lung epithelium. May play
CC important roles in developing neural, gastrointestinal and
CC cardiovascular tissues.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- TISSUE SPECIFICITY: Highest expression in lung. Lower expression
CC in spleen, skeletal muscle, brain, kidney and small intestine.
CC -!- DEVELOPMENTAL STAGE: Expressed in developing lung (only distal
CC epithelium), neural, intestinal and cardiovascular tissues.
CC -!- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 C2H2-TYPE ZINC FINGER.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; AF339106; AAK69651.1; -.
DR MGD; MGI:2148705; Foxp2.
DR PROSITE; PS00657; FORK_HEAD_1; FALSE_NEG.
DR PROSITE; PS00658; FORK_HEAD_2; FALSE_NEG.
DR PROSITE; PS50039; FORK_HEAD_3; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; FALSE_NEG.
KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein.

FT DNA_BIND 503 593 FORK-HEAD.
FT ZN_FING 345 370 C2H2-TYPE.
FT DOMAIN 53 56 POLY-GLN.
FT DOMAIN 123 126 POLY-GLN.
FT DOMAIN 131 136 POLY-GLN.
FT DOMAIN 152 191 POLY-GLN.
FT DOMAIN 200 208 POLY-GLN.
FT DOMAIN 222 230 POLY-GLN.
SQ SEQUENCE 714 AA; 79820 MW; BCDFB80E28398609 CRC64;

alignment_scores:

Quality: 198.50 Length: 352
Ratio: 1.196 Gaps: 14
Percent Similarity: 47.159 Percent Identity: 25.284

alignment_block:

US-09-664-641-10 x FXP2_MOUSE ..

Align seg 1/1 to: FXP2_MOUSE from: 1 to: 714

```
850 GAAGAGGAAGAGGAGGAGGAAGTAGAAAATGAGGAACAAGATTCTCAGAA 899
  3 GlnGluSerValThrGluThrIleSerAsnSerSerMet...AsnGlnAs 18

900 TGAGGGTAGTACAGATGAGAAGTCAAGCCCTGCCAGCTCTCAAGAAGGGT 949
 18 nGlyMetSerThrLeuSerSerGlnLeuAspAlaGlySerArgAspGlyA 35

950 CTCCTTCAGGTGACCAGCAGTTTTCACCTAAATCCAACACTGAAAAATCT 999
 35 rgSerSerGlyAsp.....ThrSerSerGluValSer 45

1000 AAAGGGGAATTAATGTTTGATGATTCTTCAGATTCATCACCGGAAAAACA 1049
  46 ThrValGluLeuLeuHisLeuGlnGlnGlnGlnAla..... 57

1050 GGAGAGAAATTTAAACTGGACCCCGGCCGAAGTCCCACAGTTAGCTGCAG 1099
  58 .....LeuGlnAlaA 61

1100 CAAAACGCAGGCTGCCTCAGGGAAAGGAGCCTGGGTTGATTAACTTGTGT 1149
  61 laArgGlnLeuLeuLeuGlnGlnGlnThrSerGlyLeu..... 73

1150 GCCAATGTCCCACCCGTCCCAGGTAACATTTTGCCCCCTGAGGTCCGGGG 1199
  73 ..... 73

1200 TAATTTAATGGCTGCTGGACAAAACCTCCAAAGTTCTGAAAGATCAGAA. 1248
  74 .....LysSerProLysSerSerGluLysGlnArgP 84

1249 .....ATGATAGCTACCTGGAGTCCAGCTGTACGGACA 1281
  84 roLeuGlnValProValSerValAlaMetMetThrProGlnValIleThr 100

1282 CTGAGGAATATTACTAATAATGCTGACATTCAGCAGATGAACCGGCCATC 1331
 101 .....ProGl 102

1332 AAATGTAGCACATATCTTACAGACTCTTTCAGCACCTACGAAAAATTTAG 1381
 102 nGlnMetGlnGlnIleLeuGlnGlnGlnValLeuSerProGlnGlnLeuG 119

1382 AACAGCAGGTGAATCACAGCCAGCAGGGACATACAAATGCCAATGCAGTG 1431
 119 lnAlaLeuLeuGlnGlnGlnGln.....AlaVal 128

1432 CTGTTTAGCCAA.....GTGAAAGTGACTCCAGA 1460
 129 MetLeuGlnGlnGlnGlnLeuGlnGluPheTyrLysLysGlnGlnGluGl 145
```

```
 1461 GACACAC......ATGCTACAGCAGCAGCAGCAGGCCCAGCAGCAGCAGC 1504
      :   |||      :::||||||||||||||||||   ||||||||||||
  145 nLeuHisLeuGlnLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnG 162

 1505 AGCAGCACCCGGTTTTACACCTTCAGCCCCAGCAGATAATGCAGCTCCAG 1554
      |||||:::                 |||   ||||||   :::|||   |||
  162 lnGlnGlnGln.............GlnGlnGlnGlnGlnGlnGlnGlnGln 174

 1555 CAGCAGCAGCAGCAGCAGATCTCTCAGCAACCTTACCCCCAGCAGCCGCC 1604
      ||||||||||||||||||   :::||||||         ||||||
  175 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGl 191

 1605 GCATCCATTTTCACAG.........CAGCAGCAGCAGCAACAGCAGCCAC 1645
       ||||||   :::|||         ||||||||||||||||||||
  191 nHisProGlyLysGlnAlaLysGluGlnGlnGlnGlnGlnGlnGlnGlnG 208

 1646 CACCATCGCCTCAGCAG............................CAT 1665
           :::   ||||||                             :::
  208 lnLeuAlaAlaGlnGlnLeuValPheGlnGlnGlnLeuLeuGlnMetGln 224

 1666 CAGCTTTTTGGACATGATCCAGCAGTGGAGATTCCAGAAGAAGGCTTCTT 1715
      ||||||      ::::::      :::::::::   ::::::|||:::::
  225 GlnLeuGlnGlnGlnGlnHisLeuLeuSerLeuGlnArgGlnGlyLeuIl 241

 1716 ATTG............GGATGTGTGTTTGCAATTGCGGATTATCCAGAGC 1753
      :                 :::::::::   :::   :::   |||:::
  241 eSerIleProProGlyGlnAlaAlaLeuProValGlnSerLeuProGlnA 258

 1754 AG...ATGTCTGATAAGCAACTGCTGGCCACCTGGAAAAGGATAATCCAG 1800
           :::|||      ::::::          ||||||::::::
  258 laGlyLeuSerProAlaGluIleGlnGlnLeuTrpLysGluValThrGly 274

 1801 GCACAT 1806
      :::|||
  275 ValHis 276

seq_name: SwissProt_40:FXP2_HUMAN

seq_documentation_block:
ID   FXP2_HUMAN     STANDARD;      PRT;   715 AA.
AC   O15409;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Forkhead box protein P2 (CAG repeat protein 44) (Trinucleotide repeat-
DE   containing gene 10 protein).
GN   FOXP2 OR CAGH44 OR TNRC10.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND VARIANT SPCH1 HIS-553.
RX   MEDLINE=21470412; PubMed=11586359;
RA   Lai C.S.L., Fisher S.E., Hurst J.A., Vargha-Khadem F., Monaco A.P.;
RT   "A forkhead-domain gene is mutated in a severe speech and language
RT   disorder.";
RL   Nature 413:519-523(2001).
RN   [2]
RP   SEQUENCE OF 1-304 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=97369492; PubMed=9225980;
RA   Margolis R.L., Abraham M.R., Gatchell S.B., Li S.H., Kidwai A.S.,
RA   Breschel T.S., Stine O.C., Callahan C., Mcinnis M.G., Ross C.A.;
RT   "cDNAs with long CAG trinucleotide repeats from human brain.";
RL   Hum. Genet. 100:114-122(1997).
RN   [3]
RP   SEQUENCE OF 1-86 FROM N.A.
RA   Minx P., Hinds K., Sutterer C., Becker M., Ozersky P.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Transcriptional repressor that plays an important role
CC       in the specification and differentiation of lung epithelium. May
CC       play important roles in developing neural, gastrointestinal and
CC       cardiovascular tissues. Involved in neural mechanisms mediating
CC       the development of speech and language.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- ALTERNATIVE PRODUCTS: 3 isoforms; 1/I (shown here), 2/II and
CC       3/III/IV; are produced by alternative splicing.
CC   -!- TISSUE SPECIFICITY: Expressed at high levels in embryonic and
CC       adult lung.
CC   -!- DISEASE: Defects in FOXP2 are the cause of an autosomal dominant
CC       speech and language disorder with orofacial dyspraxia, also known
CC       as speech-language disorder 1 (SPCH1). Affected individuals have a
CC       severe impairment in the selection and sequencing of fine
CC       orofacial movements, which are necessary for articulation. They
CC       also show deficits in several facets of language processing (such
CC       as the ability to break up words into their constituent phoneme)
CC       and grammatical skills.
CC   -!- DISEASE: Disruption of FOXP2 by a chromosomal translocation
CC       t(5;7)(q22;q31.2) is the cause of severe speech and language
CC       impairment.
CC   -!- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 C2H2-TYPE ZINC FINGER.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF337817; AAL10762.1; -.
DR   EMBL; U80741; AAB91439.1; -.
DR   EMBL; AC003992; AAB96326.1; -.
DR   MIM; 605317; -.
DR   MIM; 602081; -.
DR   PROSITE; PS00657; FORK_HEAD_1; FALSE_NEG.
DR   PROSITE; PS00658; FORK_HEAD_2; FALSE_NEG.
DR   PROSITE; PS50039; FORK_HEAD_3; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; FALSE_NEG.
KW   Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW   Nuclear protein; Chromosomal translocation; Disease mutation;
KW   Alternative splicing.
FT   DNA_BIND     504     594       FORK-HEAD.
FT   ZN_FING      346     371       C2H2-TYPE.
FT   DOMAIN        53      56       POLY-GLN.
FT   DOMAIN       123     126       POLY-GLN.
FT   DOMAIN       131     136       POLY-GLN.
FT   DOMAIN       152     191       POLY-GLN.
FT   DOMAIN       200     209       POLY-GLN.
FT   DOMAIN       223     231       POLY-GLN.
FT   VARSPLIC       1      92       MISSING (IN ISOFORM 3).
FT   VARIANT      553     553       R -> H (IN SPCH1).
FT                                  /FTId=VAR_012278.
FT   CONFLICT     134     134       Q -> H (IN REF. 2).
FT   CONFLICT     290     304       DLTTNNSSSTTSSNT -> EEFPVQGPAAVCAGL (IN
FT                                  REF. 2).
SQ   SEQUENCE   715 AA;  79919 MW;  4F9FBDB6D90516E0 CRC64;

alignment_scores:
            Quality:  196.00              Length:    353
              Ratio:  1.181                 Gaps:     14
 Percent Similarity:  47.025   Percent Identity: 24.929

alignment_block:
 US-09-664-641-10 x FXP2_HUMAN   ..

 Align seg 1/1  to: FXP2_HUMAN  from: 1  to: 715

  850 GAAGAGGAAGAGGAGGAGGAAGTAGAAAATGAGGAACAAGATTCTCAGAA 899
      :::|||:::      |||   ::::::|||:::::::::   :::|||||
    3 GlnGluSerAlaThrGluThrIleSerAsnSerSerMet...AsnGlnAs 18
```

```
 900 TGAGGGTAGTACAGATGAGAAGTCAAGCCCTGCCAGCTCTCAAGAAGGGT 949
     |      ||||||   :::::::::      |||:::|||:::::::|||
  18 nGlyMetSerThrLeuSerSerGlnLeuAspAlaGlySerArgAspGlyA 35

 950 CTCCTTCAGGTGACCAGCAGTTTTCACCTAAATCCAACACTGAAAAATCT 999
       |||||||||                  ::::::::::|||   |||
  35 rgSerSerGlyAsp.................ThrSerSerGluValSer 45

1000 AAAGGGGAATTAATGTTTGATGATTCTTCAGATTCATCACCGGAAAAACA 1049
       ||||||:::     ::::::::::::::
  46 ThrValGluLeuLeuHisLeuGlnGlnGlnGlnAla.............. 57

1050 GGAGAGAAATTTAAACTGGACCCCGGCCGAAGTCCCACAGTTAGCTGCAG 1099
                                          |||   ||||
  58 .....................................LeuGlnAlaA 61

1100 CAAAACGCAGGCTGCCTCAGGGAAAGGAGCCTGGGTTGATTAACTTGTGT 1149
     ||:::::   |||   |||   :::     ||||||
  61 laArgGlnLeuLeuLeuGlnGlnGlnThrSerGlyLeu............ 73

1150 GCCAATGTCCCACCCGTCCCAGGTAACATTTTGCCCCCTGAGGTCCGGGG 1199

  73 ............................................... 73

1200 TAATTTAATGGCTGCTGGACAAAACCTCCAAAGTTCTGAAAGATCAGAA. 1248
                       ::::::   :::||||||:::::::::::::
  74 ..................LysSerProLysSerSerAspLysGlnArgP 84

1249 ..................ATGATAGCTACCTGGAGTCCAGCTGTACGGACA 1281
                        :::|||      :::|||   |||   |||
  84 roLeuGlnValProValSerValAlaMetMetThrProGlnValIleThr 100

1282 CTGAGGAATATTACTAATAATGCTGACATTCAGCAGATGAACCGGCCATC 1331
                                                 |||::
 101 ............................................ProGl 102

1332 AAATGTAGCACATATCTTACAGACTCTTTCAGCACCTACGAAAAATTTAG 1381
     :::::::   :::|||||||||                  ::::::|||:
 102 nGlnMetGlnGlnIleLeuGlnGlnGlnValLeuSerProGlnGlnLeuG 119

1382 AACAGCAGGTGAATCACAGCCAGCAGGGACATACAAATGCCAATGCAGTG 1431
     ::      ::::::::::::|||                     ||||||
 119 lnAlaLeuLeuGlnGlnGlnGln.....................AlaVal 128

1432 CTGTTTAGCCAA.....................GTGAAAGTGACTCCAGA 1460
     :::::::::|||                    |||         ::
 129 MetLeuGlnGlnGlnGlnLeuGlnGluPheTyrLysLysGlnGlnGluGl 145

1461 GACACAC......ATGCTACAGCAGCAGCAGCAGGCCCAGCAGCAGCAGC 1504
     :   |||      :::||||||||||||||||||   ||||||||||||
 145 nLeuHisLeuGlnLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnG 162

1505 AGCAGCACCCGGTTTTACACCTTCAGCCCCAGCAGATAATGCAGCTCCAG 1554
     |||||:::                 |||   ||||||   :::|||   |||
 162 lnGlnGlnGln............GlnGlnGlnGlnGlnGlnGlnGln 174

1555 CAGCAGCAGCAGCAGCAGATCTCTCAGCAACCTTACCCCCAGCAGCCGCC 1604
     ||||||||||||||||||   :::||||||         ||||||
 175 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGl 191

1605 GCATCCA............TTTTCACAGCAGCAGCAGCAGCAACAGCAGC 1642
       ||||||             :::||||||||||||||||||||||||
 191 nHisProGlyLysGlnAlaLysGluGlnGlnGlnGlnGlnGlnGlnG 208

1643 CACCACCATCGCCTCAGCAG........................... 1662
           :::   ||||||
 208 lnGlnLeuAlaAlaGlnGlnLeuValPheGlnGlnGlnLeuLeuGlnMet 224

1663 CATCAGCTTTTTGGACATGATCCAGCAGTGGAGATTCCAGAAGAAGGCTT 1712
     :::|||||      ::::::      :::::::::   :::::::|||::
 225 GlnGlnLeuGlnGlnGlnGlnHisLeuLeuSerLeuGlnArgGlnGlyLe 241

1713 CTTATTG............GGATGTGTGTTTGCAATTGCGGATTATCCAG 1750

     ::::                    :::::::::   :::   :::   |||:
  241 uIleSerIleProProGlyGlnAlaAlaLeuProValGlnSerLeuProG 258

     1751 AGCAG...ATGTCTGATAAGCAACTGCTGGCCACCTGGAAAAGGATAATC 1797
          ::      :::|||      ::::::      ||||||::::::
  258 lnAlaGlyLeuSerProAlaGluIleGlnGlnLeuTrpLysGluValThr 274

     1798 CAGGCACAT 1806
          :::|||
  275 GlyValHis 277
```

seq_name: SwissProt_40:MCM1_YEAST

seq_documentation_block:

```
ID   MCM1_YEAST     STANDARD;      PRT;   286 AA.
AC   P11746;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Pheromone receptor transcription factor (GRM/PRTF protein).
GN   MCM1 OR FUN80 OR YMR043W OR YM9532.08.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90249735; PubMed=2159934;
RA   Ammerer G.;
RT   "Identification, purification, and cloning of a polypeptide
RT   (PRTF/GRM) that binds to mating-specific promoter elements in
RT   yeast.";
RL   Genes Dev. 4:299-312(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89141759; PubMed=3066908;
RA   Passmore S., Maine G.T., Elble R., Christ C., Tye B.K.;
RT   "Saccharomyces cerevisiae protein involved in plasmid maintenance is
RT   necessary for mating of MAT alpha cells.";
RL   J. Mol. Biol. 204:593-606(1988).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88030692; PubMed=3311883;
RA   Dubois E., Bercy J., Descamps F., Messenguy F.;
RT   "Characterization of two new genes essential for vegetative growth in
RT   Saccharomyces cerevisiae: nucleotide sequence determination and
RT   chromosome mapping.";
RL   Gene 55:265-275(1987).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Odell C., Bowman S., Barrell B.G., Rajandream M.A.;
RL   Submitted (FEB-1995) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (2.25 ANGSTROMS) OF COMPLEX WITH MAT2.
RX   MEDLINE=98140702; PubMed=9490409;
RA   Tan S., Richmond T.J.;
RT   "Crystal structure of the yeast MATalpha2/MCM1/DNA ternary complex.";
RL   Nature 391:660-666(1998).
CC   -!- FUNCTION: INTERACTS WITH THE ALPHA-2 REPRESSOR OR WITH THE
CC       ALPHA-1 ACTIVATOR THEREBY REGULATING THE EXPRESSION OF MATING-
CC       TYPE-SPECIFIC GENES.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: BELONGS TO THE MADS DOMAIN FAMILY OF TRANSCRIPTION
CC       FACTORS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
CC   ---------------------------------------------------------------------
DR   EMBL; X52453; CAA36691.1; -.
DR   EMBL; X14187; CAA32389.1; -.
DR   EMBL; M17511; AAA34609.1; -.
DR   EMBL; Z48502; CAA88409.1; -.
DR   PIR; A34599; A34599.
DR   PDB; 1MNM; 18-MAR-98.
DR   TRANSFAC; T00500; -.
DR   TRANSFAC; T00501; -.
DR   SGD; S0004646; MCM1.
DR   InterPro; IPR002100; MADS-box.
DR   Pfam; PF00319; SRF-TF; 1.
DR   PRINTS; PR00404; MADSDOMAIN.
DR   SMART; SM00432; MADS; 1.
DR   PROSITE; PS00350; MADS_BOX_1; 1.
DR   PROSITE; PS50066; MADS_BOX_2; 1.
KW   Transcription regulation; DNA-binding; Nuclear protein; 3D-structure.
FT   DOMAIN       18     72       MADS.
FT   DOMAIN       98    120       ASP/GLU-RICH (ACIDIC).
FT   CONFLICT      9      9       P -> S (IN REF. 3).
FT   CONFLICT     37     37       S -> F (IN REF. 3).
FT   CONFLICT    156    157       GA -> AR (IN REF. 3).
FT   CONFLICT    158    286       MISSING (IN REF. 3).
SQ   SEQUENCE   286 AA;  32802 MW;  FD75B4BF549E9E3B CRC64;


alignment_scores:
            Quality:  192.00            Length:    350
              Ratio:  1.324               Gaps:     12
 Percent Similarity:  41.429   Percent Identity: 23.143

alignment_block:
 US-09-664-641-10 x MCM1_YEAST   ..

 Align seq 1/1  to: MCM1_YEAST  from: 1  to: 286

     667 AATAAGAAATGCACGCATTTGATTGTTCCAGAGCCAAAGGGGGGGAAATA 716
         ||||||         |||:::         :::   |||   |||
      28 AsnLysThrArgArgHisValThrPheSerLysArgLysHisGly..... 42

     717 CGAATGTGCTTTAAAGCGAGCA...AGTATTAAAATTGTGACTCCTGACT 763
                 :::|||:::|||   ::::::::::::::|||
      43 .......IleMetLysLysAlaPheGluLeuSerValLeuThrGlyThrG 57

     764 GGGTTCTGGATTGCGTATCAGAGAAAACCAAAAAGGACGAAGCATTTTAT 813
         ||||||      |||   :::::::|||             :::|||
      57 lnValLeuLeuLeuValValSerGluThrGlyLeuValTyrThrPheSer 73

     814 CATCCTCGTCTG......ATTATTTATGAAGAGCAA.............. 843
         |||::::::      |||:::   ::::::|||
      74 ThrProLysPheGluProIleValThrGlnGlnGluGlyArgAsnLeuIl 90

     844 .....GAAGAGGAAGAGGAAGAGGAGGAGGAAG 871
                        ::::::||||||||||:::||||||||||
      90 eGlnAlaCysLeuAsnAlaProAspAspGluGluGluAspGluGluGluA 107

     872 TAGAAAATGAGGAACAAGATTCTCAGAATGAGGGTAGTACAGATGAGAAG 921
         ::::::::::::|||:::::::::::::|||:::      ::::::
     107 spGlyAspAspAspAspAspAspAspAspGlyAsnAspMetGlnArg 123

     922 TCAAGCCCTGCCAGCTCTCAAGAAGGGTCTCCTTCAGGTGACCAGCAGTT 971
         ::::::|||   ::::::|||
     124 GlnGlnProGlnGlnGlnGln..... ..... 130

     972 TTCACCTAAATCCAACACTGAAAAATCTAAAGGGGAATTAATGTTTGATG 1021

     130 ..................................................  130

    1022 ATTCTTCAGATTCATCACCGGAAAAACAGGAGAGAAATTTAAACTGGACC 1071
                        |||::::::|||::::::   ||||||
     131 ................ProGlnGlnGlnGlnGlnValLeuAsnAlaHis 141

    1072 CCGGCCGAAGTCCCACAGTTAGCTGCAGCAAAACGCAGGCTGCCTCAGGG 1121
```

```
                ::::::   :::|||
     142 AlaAsnSerLeuGlyHisLeu..... 148

    1122 AAAGGAGCCTGGGTTGATTAACTTGTGTGCCAATGTCCCACCCGTCCCAG 1171
                                        |||          ||||||:
     149 ..........................AsnGlnAspGlnValProA 155

    1172 GTAACATTTTGCCCCCTGAGGTCCGGGGTAATTTAATGGCTGCTGGACAA 1221
         :::::   |||      ||||||:::::::::|||:::::::::::::
     155 laGlyAlaLeuLysGlnGluValLysSerGlnLeuLeuGlyGlyAla... 170

    1222 AACCTCCAAAGTTCTGAAAGATCAGAAATGATAGCTACCTGGAGTCCAGC 1271

     170 ..................................................  170

    1272 TGTACGGACACTGAGGAATATTACTAATAATGCTGACATTCAGCAGATGA 1321
                        |||   ::::::|||:::   |||||||||
     171 ..............AsnProAsnGlnAsnSerMetIleGlnGln.... 180

    1322 ACCGGCCATCAAATGTAGCACATATCTTACAGACTCTTTCAGCACCTACG 1371

     180 ..................................................  180

    1372 AAAAATTTAGAACAGCAGGTGAATCACAGCCAGCAGGGACATACAAATGC 1421
                                        ||||||   ||||||
     181 ..........................GlnGlnHisHisThr..... 185

    1422 CAATGCAGTGCTGTTTAGCCAAGTGAAAGTGACTCCAGAGACACACATGC 1471

     185 ..................................................  185

    1472 TACAGCAGCAGCAGCAGGCCCAGCAGCAGCAGCAGCAGCACCCGGTTTTA 1521
              |||::::::|||   ||||||||||||||||||
     186 .....GlnGlnSerGlnProGlnGlnGlnGlnGlnGln..... 196

    1522 CACCTTCAGCCCCAGCAGATAATGCAGCTCCAGCAGCAGCAGCAGCAGCA 1571
              |||||||||               ||||||::::::||||||||
     197 ......GlnProGln..............GlnGlnMetSerGlnGlnGl 206

    1572 GATCTCTCAGCAACCTTACCCCCAGCAGCCG...CCGCATCCATTTTCAC 1618
         |:::|||||:::|||   |||||||||      |||||||||   ::::
     206 nMetSerGlnHisProArgProGlnGlnGlyIleProHisProGlnGlnS 223

    1619 AGCAGCAGCAGCAGCAACAGCAGCCACCACCATCGCCTCAGCAGCATCAG 1668
         ::|||   ||||||||||||||||         :::   ||||||:::|||
     223 erGlnProGlnGlnGlnGlnGlnGlnGlnGlnGlnLeuGlnGlnGlnGln 239

seq_name: SwissProt_40:KNQ1_YEAST

seq_documentation_block:
ID   KNQ1_YEAST     STANDARD;      PRT;   756 AA.
AC   P53894;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Probable serine/threonine-protein kinase YNL161W (EC 2.7.1.-).
GN   YNL161W OR N1727.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / FY1679;
RX   MEDLINE=96287653; PubMed=8686380;
RA   Nasr F., Becam A.-M., Herbert C.J.;
RT   "The sequence of 36.8 kb from the left arm of chromosome XIV reveals
RT   24 complete open reading frames: 18 correspond to new genes, one of
RT   which encodes a protein similar to the human myotonic dystrophy
RT   kinase.";
RL   Yeast 12:169-175(1996).
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC       SIMILAR TO N.CRASSA PROTEIN KINASE COT1.
```

```
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X92517; CAA63278.1; -.
DR   EMBL; Z71437; CAA96048.1; -.
DR   HSSP; P05132; 1BKX.
DR   SGD; S0005105; YNL161W.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR000961; Pkinase_C.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00433; pkinase_C; 1.
DR   SMART; SM00133; S_TK_X; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Hypothetical protein; Transferase; Serine/threonine-protein kinase;
KW   ATP-binding.
FT   DOMAIN       23     34       POLY-GLN.
FT   DOMAIN      160    164       POLY-SER.
FT   DOMAIN      214    232       POLY-GLN.
FT   DOMAIN      235    250       POLY-GLN.
FT   DOMAIN      352    672       PROTEIN KINASE.
FT   NP_BIND     358    366       ATP (BY SIMILARITY).
FT   BINDING     381    381       ATP (BY SIMILARITY).
FT   ACT_SITE    475    475       BY SIMILARITY.
SQ   SEQUENCE   756 AA;  86946 MW;  87EBCD2C3C96EE11 CRC64;


alignment_scores:
           Quality:  189.50            Length:    803
             Ratio:  0.562              Gaps:     37
 Percent Similarity:  41.968   Percent Identity: 21.669

alignment_block:
 US-09-664-641-10 x KNQ1_YEAST   ..

 Align seg 1/1  to: KNQ1_YEAST  from: 1  to: 756

     835 CAAGAGGAAGAAGAGGAAGAGGAAGAGGAGGAGGAAGTAGAAAATGAGGA 884
         :::::::::::::::::::::::::::::::::::      ||||||
      23 GlnGlnGlnAspGlnGlnHisGlnGlnGlnGlnGlnTyrAlaAsnGluMe 39

     885 ACAAGATTCTCAGAATGAGGGTAGTACAGATGAGAAGTCAAGCCCTGCCA 934
          :::     |||                          :::   |||
      39 tAsnProTyrGln.....................GlnIleProArgP 48

     935 GCTCTCAAGAAGGGTCTCCTTCAGGTGACCAGCAGTTTTCACCTAAATCC 984
                 |||                    ||||||     |||
      48 roProAlaAlaGly.....................PheSer......Ser 55

     985 AACACTGAAAAATCTAAAGGGGAATTAATGTTTGATGATTCTTCAGATTC 1034
         |||      |||::::::|||:::           :::|||   :::::
      56 AsnTyrMetLysGluGlnGlySer.........HisGlnSerLeuGlnGl 69

    1035 ATCACCGGAAAAACAGGAGAGAAATTTAAACTGGACCCCGGCCGAAGTCC 1084
         :      :::::::::      |||||||:::         ::::::||||
      69 uHisLeuGlnArgGluThrGlyAsnLeuGlySerGlyPheThrAspValP 86

    1085 CACAGTTAGCTGCAGCAAAACGCAGGCTGCCTCAGGGAAAGGAGCCTGGG 1134
         ||   |||
      86 roAlaLeu..................................... 88

    1135 TTGATTAACTTGTGTGCCAATGTCCCACCCGTCCCAGGTAACATTTTGCC 1184
                        |||   |||   :::|||           ||
      89 .................AsnTyrProAlaThrPro...........Pr 95

    1185 CCCTGAGGTCCGGGGTAATTTAATGGCTGCTGGACAAAACCTCCAAAGTT 1234
         ||||      ::::::|||      |||::::::|||
      95 oPro......HisAsnAsnTyrAlaAlaSerAsnGln..... 105

    1235 CTGAAAGATCAGAAATGATAGCTACCTGGAGTCCAGCTGTACGGACACTG 1284
                       ||||||   |||      |||::::::      |||
     106 .............MetIleAsnThrProProProSerMetGlyGlyLeu 117

    1285 AGGAATATTACTAATAATGCTGAC......ATTCAGCAGATGAACCGGCC 1328
         :::   :::|||||||::::::      :::|||:::   |||
     118 TyrArgHisAsnAsnAsnSerGlnSerMetValGlnAsnGlyAsnGlySe 134

    1329 ATCAAATGTAGCACATATCTTACAGACTCTTTCAGCACCTACGAAAAATT 1378
          :::|||:::          |||      ||||||            ::::
     134 rGlyAsnAlaGln......LeuProGlnLeuSerProGlyGlnTyrSerI 149

    1379 TAGAACAGCAGGTGAATCACAGCCAGCAGGGACATACAAATGCCAATGCA 1428
         ::|||::::::   |||:::::::   :::|||   ::::::::::::
     149 leGluSerGluTyrAsnGlnAsnLeuAsnGlySerSerSerSerSerPro 165

    1429 GTGCTGTTTAGCCAAGTG..............................  1446
                    |||:::
     166 PheHisGlnProGlnThrLeuArgSerAsnGlySerTyrSerSerGlyLe 182

    1447 ......................AAAGTGACTCCAGAGACACACATGCTAC 1474
                                ::::::      |||   :::   :::|
     182 uArgSerValLysSerPheGlnArgLeuGlnGlnGluGlnGluAsnValG 199

    1475 AGCAGCAGCAGCAGGCCCAGCAGCAGCAGCAGCAGCACCCGGTTTTACAC 1524
         ||   |||||||||   :::|||   ||||||||||:::
     199 lnValGlnGlnGlnLeuSerGlnAlaGlnGlnGlnAsn.........Ser 212

    1525 CTTCAGCCCCAGCAGATAATGCAGCTCCAGCAGCAGCAGCAGCAGCAGAT 1574
            |||   ||||||   :::|||   ||||||||||||||||||||
     213 ArgGlnGlnGlnGlnGlnLeuGlnTyrGlnGlnGlnGlnGlnGlnGlnGl 229

    1575 CTCTCAGCAACCTTACCCCCAGCAGCCGCCGCATCCATTTTCACAGCAGC 1624
          :::||||||            |||                  :::|||||||
     229 nGlnGlnGlnHisMetGlnIleGlnGln.............GlnGlnGlnG 242

    1625 AGCAGCAGCAACAGCAGCCACCACCATCGCCTCAGCAGCATCAGCTTTTT 1674
         ||||||||||||||||         ||||||   |||
     242 lnGlnGlnGlnGlnGlnSerGlnSerProValGln............. 254

    1675 GGACATGATCCAGCAGTGGAGATTCCAGAAGAAGGCTTCTTATTGGGATG 1724
                                        :::||||||      |||
     255 ............................SerGlyPheAsnAsnGly.. 260

    1725 TGTGTTTGCAATTGCGGATTATCCAGAGCAGATGTCTGATAAGCAACTGC 1774
               :::|||::::::|||                      :::||||
     261 .......ThrIleSerAsnTyrMetTyrPheGluArgArgProAspLeuL 275

    1775 TGGCCACCTGGAAAAGGATAATCCAGGCACATGGCGGCACTGTTGACCCC 1824
         ||:::                                 ||||||   |||
     275 euThrLys.........................GlyThrGlnAsp... 281

    1825 ACCTTCACGAGTCGATGCACGCACCTTCTCTGTGAGAGTCAAGTCAGCAG 1874
                :::   ::::::   |||      |||:::      :::||
     282 ......LysAlaAlaAlaValLysLeuLysIleGluAsnPheTyrGlnSe 296

    1875 CGCGTATGCACAGGCAATAAGAGAAAGAAAGAGATGTGTT.......... 1914
         |:::         ||||||:::::::::::|||   |||
     296 rSerValLysTyrAlaIleGluArgAsnGluArgArgValGluLeuGluT 313

    1915 ........ACTGCACACTGGTTAAACACAGTCTTAAAAAAAAAAAAAATG 1956
                ||||::|||      :::         |||:::::::::::::
     313 hrGluLeuThrSerHisAsnTrpSerGluGluArgLysSerArgGlnLeu 329

    1957 GTACCGCCGCACCGAGCCCTTCACTTCCCAGTGGCCTTCCCACCAGGAGG 2006
                                                        ||
     330 ..................................SerSerLeuGl 333
```

```
2007 AAAGCCATGTTCACAGCATATTATTTCTGTGACTGGATTTGTTGATAGTG 2056
     ||||      ||||||                        |||:::
 333 yLysLysGluSerGln.....................PheLeuArgLeuA 343

2057 ACAGAGATGACCTAAAATTAATGGCTTATTTGGCAGGTGCCAAATATACG 2106
       |||      |||:::|||      :::   :::   :::|||
 343 rgArgThrArgLeuSerLeuGluAspPheHisThr...ValLysValIle 358

2107 GGTTATCTATGCCGCAGCAACACAGTCCTCATCTGTAAAGAACCAACTGG 2156
     |||    :::   :::::::::   |||:::   |||:::   |||||
 359 GlyLysGlyAlaPheGlyGluValArgLeuValGlnLysLysAspThrGl 375

2157 T.......................TTAAAGTATGAAAAAGCCAAAGAGT 2182
     |                        ||||||   |||      |||:::
 375 yLysIleTyrAlaMetLysThrLeuLeuLysSerGluMetTyrLysLysA 392

2183 GGAGGATACCCTGTGTCAACGCCCAGTGGCTTGGCGACATTCTTCTGGGA 2232
       ::::::      |||:::|||:::          |||:::|||   |||
 392 spGlnLeuAlaHisValLysAlaGlu......ArgAspValLeuAlaGly 406

2233 AAC.......................TTTGAGGCACTGAGGCAGATTCA 2258
     :::                        :::   ::::::::::::   ||
 407 SerAspSerProTrpValValSerLeuTyrTyrSerPheGlnAspAlaGl 423

2259 GTATAGTCGCTACACGGCATTCAGTCTGCAGGATCCATTTGCCCCTACCC 2308
     ||||                :::   |||         |||   |||
 423 nTyr..............LeuTyrLeuIleMetGluPheLeuProGlyG 435

2309 AGCATTTAGTTTTAAATCTTTTAGATGCTTGGAGA.............. 2343
          |||   :::::::::|||      |||:::
 435 lyAspLeu...MetThrMetLeuIleArgTrpGlnLeuPheThrGluAsp 450

2344 GTTCCCTTAAAAGTGTCTGCAGAGTTGTTGATGAGTATAAGACTACCTCC 2393
     |||                |||||   ::::::::::|||:::
 451 ValThrArgPheTyrMetAlaGluCysIleLeuAlaIleGluThrIleHi 467

2394 CAAACTG.......................AAACAGAATGAAGTAGCTA 2419
       ||||||                        :::|||   :::   :
 467 sLysLeuGlyPheIleHisArgAspIleLysProAspAsnIleLeuIleA 484

2420 ATGTCCACCCTTCTTCCCAAAGAGCCAGAATTGAAGACGTACCACCTCCC 2469
     :::::                |||:::::::|||
 484 spIle..............ArgGlyHisIle..................... 489

2470 ACTAAAAAGCTAACTCCAGAATTGACCCCTTTTGTGCTTTTCACTGGATT 2519
                        :::|||:::   |||   |||   ||||||||
 490 .................LysLeuSerAspPheGlyLeuSerThrGlyPh 500

2520 CGAGCCTGTCCAGGTTCAACAGTATATTAAGAAGCTCTAC.......... 2559
     |:::   ::::::   ::::::|||   |||||||||
 500 cHisLysThrHisAspSerAsnTyrTyrLysLysLeuLeuGlnGlnAspG 517

2560 ............................ATTCTTGGTGGAGAGGTTGCG 2580
                                 :::::::::::::::
 517 luAlaThrAsnGlyIleSerLysProGlyThrTyrAsnAlaAsnThrThr 533

2581 GACTCTGCACAGAAGTGCACACACCTCATTGCCAGCAAAGTGACTCGCAC 2630
     ::::::|||:::|||         ||||||||||||||||||   ||
 534 AspThrAlaAsnLysArgGlnThrMetValValAspSerIleSerLeuTh 550

2631 CGTG.................................AAGTTCCTGA 2644
     |:::                                 :::   |||
 550 rMetSerAsnArgGlnGlnIleGlnThrTrpArgLysSerArgArgLeuM 567

2645 CGGCGATTTCTGTCGTGAAGCACATAGTGACGCCAGAGTGGCTG...GAA 2691
       |||   |||:::|||           ||||||::::::::
 567 etAlaTyrSerThrValGly.........ThrProAspTyrIleAlaPro 580

2692 GAATGCTTCAGGTGTCAGAAGTTCATTGATGAGCAGAACTACATTCTCCG 2741
     |||   |||      |||   :::   :::|||
 581 GluIlePheLeuTyrGlnGlyTyrGlyGlnGlu..................  591

2742 AGATGCTGAGGCAGAAGTACTTTTCTCTTTCAGCTTGGAAGAATCCTTAA 2791
```

```
 591 .....................................................  591

2792 AACGGGCACACGTTTCTCCACTCTTTAAGGCAAAATATTTTTACATCACA 2841

 591 .....................................................  591

2842 CCTGGAATCTGCCCAAGTCTTTCCACTATGAAGGCAATCGTAGAGTGTGC 2891
           |||         |||             :::   ||||||
 592 .........CysAspTrpTrpSerLeuGlyAlaIleMetTyrGluCysLe 605

2892 AGGAGGA......AAGGTGTTATCCAAGCAGCCA......TCTTTCCGGA 2929
       |||                |||:::   |||      ::::::||||
 605 uIleGlyTrpProProPheCysSerGluThrProGlnGluThrTyrArgL 622

2930 AGCTCATGGAGCACAAGCAGAACTCGAGTTTGTCGGAAATAATTTTAATA 2979
     ||:::|||:::   :::|||:::   ::::::   :::   |||   |||
 622 ysIleMetAsnPheGluGlnThrLeuGlnPheProAspAspIleHisIle 638

2980 TCCTGTGAA 2988
     |||   |||
 639 SerTyrGlu 641
```

```
seq_name: SwissProt_40:SSN6_YEAST

seq_documentation_block:
ID   SSN6_YEAST     STANDARD;      PRT;   966 AA.
AC   P14922;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   01-FEB-1995 (Rel. 31, Last annotation update)
DE   Glucose repression mediator protein.
GN   SSN6 OR CYC8 OR YBR112C OR YBR0908.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89211964; PubMed=2854095;
RA   Trumbly R.J.;
RT   "Cloning and characterization of the CYC8 gene mediating glucose
RT   repression in yeast.";
RL   Gene 73:97-111(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88065502; PubMed=3316983;
RA   Schultz J., Carlson M.;
RT   "Molecular analysis of SSN6, a gene functionally related to the SNF1
RT   protein kinase of Saccharomyces cerevisiae.";
RL   Mol. Cell. Biol. 7:3637-3645(1987).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE=92327848; PubMed=1626431;
RA   Mannhaupt G., Stucka R., Ehnle S., Vetter I., Feldmann H.;
RT   "Molecular analysis of yeast chromosome II between CMD1 and LYS2: the
RT   excision repair gene RAD16 located in this region belongs to a novel
RT   group of double-finger proteins.";
RL   Yeast 8:397-408(1992).
RN   [4]
RP   TPR REPEATS.
RX   MEDLINE=90124639; PubMed=2404612;
RA   Sikorski R.S., Boguski M.S., Goebl M., Hieter P.A.;
RT   "A repeating amino acid motif in CDC23 defines a family of proteins
RT   and a new relationship among genes required for mitosis and RNA
RT   synthesis.";
RL   Cell 60:307-317(1990).
CC   -!- FUNCTION: IT IS INVOLVED IN REPRESSION BY A1-ALPHA2 AND ALPHA2 AND
CC       IN OTHER SYSTEMS AS A GENERAL REPRESSOR OF TRANSCRIPTION. THIS
CC       PROTEIN HAS NO OBVIOUS DNA-BINDING DOMAINS. IT MIGHT NOT INTERACT
CC       DIRECTLY WITH DNA BUT WITH DNA-BOUND PROTEINS.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
```

```
CC   -!- SIMILARITY: CONTAINS 10 TPR REPEATS.
CC   -!- SIMILARITY: TO YEAST GAL1 AND CCR4.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M23440; AAA34545.1; -.
DR   EMBL; M17826; AAA35103.1; -.
DR   EMBL; X66247; CAA46973.1; -.
DR   EMBL; X78993; CAA55615.1; -.
DR   EMBL; Z35981; CAA85069.1; -.
DR   PIR; S25365; S25365.
DR   SGD; S0000316; CYC8.
DR   InterPro; IPR001440; TPR.
DR   Pfam; PF00515; TPR; 10.
DR   SMART; SM00028; TPR; 9.
KW   Transcription regulation; Repressor; Repeat; TPR repeat;
KW   Nuclear protein.
FT   DOMAIN       15     30       POLY-GLN.
FT   REPEAT       46     79       TPR 1.
FT   REPEAT       80    113       TPR 2.
FT   REPEAT      114    147       TPR 3.
FT   REPEAT      150    183       TPR 4.
FT   REPEAT      187    220       TPR 5.
FT   REPEAT      224    257       TPR 6.
FT   REPEAT      258    291       TPR 7.
FT   REPEAT      296    329       TPR 8.
FT   REPEAT      330    363       TPR 9.
FT   REPEAT      364    398       TPR 10.
FT   DOMAIN      493    556       30 X 2 AA TANDEM REPEATS OF Q-A.
FT   DOMAIN      557    587       POLY-GLN.
FT   CONFLICT    547    547       K -> Q (IN REF. 3).
SQ   SEQUENCE   966 AA;  107202 MW;  84B509CF3208C5C0 CRC64;


alignment_scores:
            Quality:  186.00              Length:    519
              Ratio:  0.805                Gaps:     24
 Percent Similarity:  44.509   Percent Identity: 23.314

alignment_block:
 US-09-664-641-10 x SSN6_YEAST   ..

 Align seg 1/1  to: SSN6_YEAST  from: 1  to: 966

    1087 CAGTTAGCTGCAGCAAAACGCAGGCTGCCTCAGGGAAAGGAGCCTGGGTT 1136
          :::|||   |||      ::::::|||         ::::::||||||
     404 ArgLeuGluAlaLeuThrLysGlnLeu.........GluAsnProGlyAs 417

    1137 GATTAACTTGTGT.............GCCAATGTCCCACCCGTCCCAGGTA 1174
          ||||||                   :::|||:::   |||:::|||
     417 nIleAsnLysSerAsnGlyAlaProThrAsnAlaSerProAlaProProP 434

    1175 ACATT..............TTGCCCCCTGAGGTCCGGGGTAATTTAATG 1209
           :::               |||   |||:::   :::||||||   :::
     434 roValIleLeuGlnProThrLeuGlnProAsnAspGlnGlyAsnProLeu 450

    1210 GCTGCTGGACAAAACCTCCAAAGTTCTGAA...AGATCAGAAATGATAGC 1256
           :::      :::   ||||||:::::      ::::::|||:::
     451 AsnThrArgIleSerAlaGlnSerAlaAsnAlaThrAlaSerMetValGl 467

    1257 TACCTGGAGTCCAGCTGTACCGACACTGAGGAATATTACTAATAATGCTG 1306
                 ||||||   :::|||
     467 nGlnGlnHisProAlaGlnGlnThr..................... 475

    1307 ACATTCAGCAGATGAACCGGCCATCAAATGTAGCACATATCTTACAGACT 1356
                         |||   |||   :::      :::   :::
     476 ....................ProIleAsnSerSerAlaThrMetTyrSer 485

    1357 CTTTCAGCACCTACGAAAAATTTAGAACAGCAGGTGAATCACAGCCAGCA 1406
           :::|||   :::   :::|||:::   |||:::::::   :::   ||
     486 AsnGlyAla...SerProGlnLeuGlnAlaGlnAlaGlnAlaGlnAlaGl 501

    1407 GGGACATACAAATGCCAATGCAGTGCTGTTTAGCCAAGTGAAAGTGACTC 1456
         |:::::::::::|||:::|||          :::|||:::::::::
     501 nAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnA 518

    1457 CAGAGACACAC...ATGCTACAGCAGCAGCAGCAGGCCCAGCAGCAGCAG 1503
           :::::::::   :::   |||   |||   |||||||||   |||
     518 laGlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAla 534

    1504 CAGCAGCACCCGGTTTTACACCTTCAGCCCCAG..................... 1536
         |||   |||          :::   |||   |||
     535 GlnAlaHisAlaGlnAlaGlnAlaGlnAlaGlnAlaLysAlaGlnAlaGl 551

    1537 ....CAGATAATGCAGCTCCAGCAGCAGCAGCAGCAGCAGATCTCTCAGC 1582
             |||   :::      ||||||||||||||||||||||   :::||||
     551 nAlaGlnAlaGlnAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnG 568

    1583 AACCTTACCCCCAGCAGCCGCCGCATCCATTTTCACAGCAGCAGCAGCAG 1632
         ||         ||||||      :::      :::|||||||||||||||
     568 lnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 584

    1633 CAACAGCAGCCACCACCATCGCCTCAGCAGCATCAGCTTTTTGGACATGA 1682
         |||||||||      |||   |||:::|||   ||||||
     585 GlnGlnGlnLeuGlnProLeuProArgGln...GlnLeuGlnGlnLysGl 600

    1683 TCCAGCAGTGGAGATT......CCAGAAGAAGGCTTCTTATTGGGATGTG 1726
           :::|||::::::      |||:::::::|||
     600 yValSerValGlnMetLeuAsnProGlnGlnGly.............. 611

    1727 TGTTTGCAATTGCGGATTATCCAGAGCAGATGTCTGATAAGCAACTGCTG 1776
                                                     |||
     612 ....................................GlnProTyr 614

    1777 GCCACCTGGAAAAGGATAATCCAGGCACATGGCGGCACTGTTGACCCCAC 1826
           |||         :::|||||||||||||          :::::::|||
     615 IleThrGlnProThrValIleGlnAlaHis......GlnLeuGlnProPh 629

    1827 CTTCACGAGTCGATGCACGCACCTTCTCTGTGAGAGTCAAGTCAGCAGCG 1876
           |||:::      |||         :::||||||:::
     629 eSerThrGlnAlaMetGluHisProGln...SerSerGlnLeuProProG 645

    1877 CGTATGCACAGGCAATAAGAGAAAGAAAGAGATGTGTTACTGCACACTGG 1926
                |||
     645 lnGlnGlnGlnLeu..................................... 649

    1927 TTAAACACAGTCTTAAAAAAAAAAAAAAATGGTACCGCCGCACCGAGCCCT 1976
                                  :::::::::      |||::::::
     650 ....................GlnSerValGlnHisProGlnGlnLeuGl 659

    1977 TCACTTCCCAGTGGCCTTCCCACCAGGAGGAAAGCCATGTTCACAGCATA 2026
               |||   |||      |||      :::|||      ||||||
     659 nGlyGlnProGlnAlaGlnAlaPro......GlnProLeuIleGlnHisA 674

    2027 TTATTTCTGTGACTGGATTTGTTGATAGTGACAGAGATGACCTAAAATTA 2076
           :::                          :::::::::   |||
     674 snVal.......................GluGlnAsnValLeuProGln 682

    2077 ATGGCTTATTTGGCAGGTGCCAAATATACGGGTTATCTATGCCGCAGCAA 2126
               |||:::   ||||||   :::||| ::::::   ||||||::::
     683 LysArgTyrMetGluGlyAlaIleHisThr.LeuValAspAlaAlaVals 699

    2127 CACAGTCCTCATCTGTAAAGAACCAACTGGTTTAAAGTATGAAAAAGCCA 2176
         :::::|||:::   ::::::|||            :::   |||:::|||
     699 erSerSerThrHisThrGluAsn............AsnThrLysSerPro 711

    2177 AAGAGTGGAGGATACCCTGTGTCAACGCCCAGTGGCTTGGCGACATTCTT 2226
         ::::::      :::   :::   |||   :::
     712 ArgGlnProThrHisAlaIleProThrGlnAla..................... 722
```

```
     2227 CTGGGAAACTTTGAGGCACTGAGGCAGATTCAGTATAGTCGCTACACGGC 2276
                                                 |||||| |||
      723 ..................................ProAlaThr.Gly 726

     2277 ATTCAGTCTGCAGGATCCATTTGCCCCTACCCAGCATTTA..... 2316
          |||   :::||| :::|||   :::      |||   |||
      727 IleThrAsnAla.GluProGlnValLysLysGlnLysLeuAsnSerProA 743

     2317 .....GTTTTAAATCTTTTAGATGCTTGGAGAGTTCCCTTAAAAGTGTCT 2361
               :::|||   |||          :::  :::::  :::
      743 snSerAsnIleAsnLysLeuValAsnThrAlaThrSerIleGluGluAsn 759

     2362 GCAGAGTTGTTGATGAGTATAAGACTACCT..... 2391
          |||:::      :::|||   :::   |||
      760 AlaLysSerGluValSerAsnGlnSerProAlaValValGluSerAsnTh 776

     2392 .....CCCAAACTGAAACAGAATGAAGTAGCTA 2419
                              :::|||   |||::::::   :
      776 rAsnAsnThrSerGlnGluGluLysProValLysAlaAsnSerIleProS 793

     2420 ATGTC..............................CAGCCTTCTTCCCAA 2439
          ::|||                            :::|||:::::::::
      793 erValIleGlyAlaGlnGluProProGlnGluAlaSerProAlaGluGlu 809

     2440 AGAGCCAGAATTGAAGACGTACCACCTCCCACTAAAAAGCTAACTCCAGA 2489
            ::::::      :::|||   |||   ||||||   |||:::  ||
      810 AlaThrLysAlaAlaSerValSerProSerThrLysProLeuAsnThrGl 826

     2490 A 2490
          |
      826 u 826
```

seq_name: SwissProt_40:GLTA_WHEAT

seq_documentation_block:

```
ID   GLTA_WHEAT     STANDARD;      PRT;   356 AA.
AC   P10385;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Glutenin, low molecular weight subunit precursor.
OS   Triticum aestivum (Wheat).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC   Triticeae; Triticum.
OX   NCBI_TaxID=4565;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. YAMHILL;
RX   MEDLINE=89083577; PubMed=3205747;
RA   Pitts E.G., Rafalski J.A., Hedgcoth C.;
RT   "Nucleotide sequence and encoded amino acid sequence of a genomic
RT   gene region for a low molecular weight glutenin.";
RL   Nucleic Acids Res. 16:11376-11376(1988).
CC   -!- FUNCTION: GLUTENINS ARE THE HIGH MOLECULAR WEIGHT SEED STORAGE
CC       PROTEIN OF WHEAT ENDOSPERM. THOUGHT TO BE RESPONSIBLE FOR THE
CC       VISCO-ELASTIC PROPERTY OF WHEAT DOUGH.
CC   -!- SUBUNIT: DISULFIDE-BRIDGE LINKED AGGREGATES.
CC   -!- MISCELLANEOUS: GLUTENINS ARE CODED BY SEVERAL GENES ON EACH OF THE
CC       GROUP 1 CHROMOSOMES OF WHEAT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X07747; CAA30570.1; -.
DR   PIR; S01992; S01992.
DR   InterPro; IPR003612; AAI.
DR   InterPro; IPR001768; Cereal_tryp_amyl_inh.
DR   InterPro; IPR001954; Glia_glutenin.
DR   Pfam; PF00234; tryp_alpha_amyl; 1.
DR   PRINTS; PR00208; GLIADGLUTEN.
DR   SMART; SM00499; AAI; 1.
KW   Seed storage protein; Repeat; Multigene family; Signal.
FT   SIGNAL       1     19       POTENTIAL.
FT   CHAIN       20    356       GLUTENIN, LOW MOLECULAR WEIGHT SUBUNIT.
SQ   SEQUENCE   356 AA;  41020 MW;  AA2613FC0DD4DCA5 CRC64;
```

alignment_scores:

```
           Quality:  185.50            Length:    124
             Ratio:  2.576               Gaps:      5
 Percent Similarity:  58.065   Percent Identity: 45.968
```

alignment_block:

US-09-664-641-10 x GLTA_WHEAT ..

Align seq 1/1 to: GLTA_WHEAT from: 1 to: 356

```
    1330 TCAAATGTAGCACATATCTTACAGACTCTTTCAGCACCTACGAAAAATTT 1379
         |||   ||||||:::|||   |||       :::|||||||      :::
      16 SerAlaValAlaGlnIleSerGlnGlnGlnGlnAlaProProPheSerGl 32

    1380 AGAACAGCAGGTGAATCACAGCCAGCAGGGACATACAAATGCCAATGCAG 1429
          :::||||||         |||||||||   :::
      32 nGlnGlnGlnProProPheSerGlnGlnGlnGlnProPro..... 45

    1430 TGCTGTTTAGCCAAGTGAAAGTGACTCCAGAGACACACATGCTACAGCAG 1479
               |||||||||   :::   :::|||          :::   ||||||
      46 .....PheSerGlnGlnGlnGlnSerPro.........PheSerGlnGln 57

    1480 CAGCAGCAGGCCCAGCAGCAGCAGCAGCAGCACCCGGTTTTACACCTTCA 1529
         |||||||||          ||||||||||:::|||   :::      ||
      58 GlnGlnGlnProProPheAlaGlnGlnGlnGlnProProPheSerGlnGl 74

    1530 GCCC............CAGCAGATAATGCAGCTCCAGCAGCAGCAGCAGC 1567
         ||||            ||||||          ::::::||||||||||||
      74 nProProIleSerGlnGlnGlnGlnProProPheSerGlnGlnGlnGlnP 91

    1568 AGCAGATCTCTCAGCAA.........CCTTACCCCCAGCAGCCGCCGCAT 1608
           |||:::|||||||||          ||||||   ||||||
      91 roGlnPheSerGlnGlnGlnGlnProProTyrSerGlnGlnGlnGlnPro 107

    1609 CCATTTTCACAGCAGCAGCAG............CAGCAACAGCAGCCACC 1646
         |||:::||||||||||||||||          ||||||||||||||||||
     108 ProTyrSerGlnGlnGlnGlnProProPheSerGlnGlnGlnGlnProPr 124

    1647 ACCATCGCCTCAGCAGCATCAG 1668
         |   |||   ||||||:::|||
     124 oPheSerGlnGlnGlnGlnGln 131
```

seq_name: SwissProt_40:BRD4_HUMAN

seq_documentation_block:

```
ID   BRD4_HUMAN     STANDARD;      PRT;  1362 AA.
AC   O60885; Q96PD3;
DT   16-OCT-2001 (Rel. 40, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Bromodomain-containing protein 4 (HUNK1 protein).
GN   BRD4 OR HUNK1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   French C.A., Fletcher J.A.;
RT   "Human BRD4 protein.";
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 1-722 FROM N.A.
```

```
RC   TISSUE=Placenta;
RA   Weber B.;
RL   Submitted (MAR-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- SIMILARITY: CONTAINS 2 BROMODOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF386649; AAL26987.1; -.
DR   EMBL; Y12059; CAA72780.1; -.
DR   InterPro; IPR001487; Bromodomain.
DR   Pfam; PF00439; bromodomain; 2.
DR   PRINTS; PR00503; BROMODOMAIN.
DR   SMART; SM00297; BROMO; 2.
DR   PROSITE; PS00633; BROMODOMAIN_1; 1.
DR   PROSITE; PS50014; BROMODOMAIN_2; 2.
KW   Bromodomain; Repeat; Nuclear protein.
FT   DOMAIN       75    147       BROMODOMAIN 1.
FT   DOMAIN      368    440       BROMODOMAIN 2.
FT   DOMAIN      535    594       LYS-RICH.
FT   DOMAIN      692    717       SER-RICH.
FT   DOMAIN      703    714       POLY-SER.
FT   DOMAIN      738    743       POLY-HIS.
FT   DOMAIN      757    761       POLY-PRO.
FT   DOMAIN      764    770       POLY-PRO.
FT   DOMAIN      771    775       POLY-GLN.
FT   DOMAIN      776    783       POLY-PRO.
FT   DOMAIN      954    964       POLY-PRO.
FT   DOMAIN      974    986       POLY-PRO.
FT   DOMAIN     1011   1014       POLY-PRO.
FT   DOMAIN     1028   1033       POLY-PRO.
FT   DOMAIN     1283   1300       POLY-GLN.
FT   DOMAIN     1301   1308       POLY-ALA.
FT   DOMAIN     1335   1338       POLY-ARG.
FT   CONFLICT    720    721       EM -> GP (IN REF. 2).
SQ   SEQUENCE   1362 AA;  152219 MW;  D52EFF1CF9960907 CRC64;


alignment_scores:
            Quality:  182.50              Length:    432
              Ratio:  1.037                 Gaps:     21
 Percent Similarity:  40.741   Percent Identity: 23.511

alignment_block:
 US-09-664-641-10 x BRD4_HUMAN   ..

 Align seg 1/1  to: BRD4_HUMAN  from: 1  to: 1362

     658 CTAACCCTCAATAAGAAATGCACGCATTTGATTGTTCCAGAGCCAAAGGG 707
         |||   :::|||||||                  :::|||      |||
     620 LeuAspIleAsnLys.................LeuProGlyGluLysLe 630

     708 GGGGAAA......TACGAATGTGCTTTAAAGCGAGCAAGTATTAAAATTG 751
          ||:::      :::           ::::::   |||:::|||
     630 uGlyArgValValHisIleIleGlnSerArgGluProSerLeuLysAsnS 647

     752 TGACTCCTGAC..................................... 762
         :::||||||
     647 erAsnProAspGluIleGluIleAspPheGluThrLeuLysProSerThr 663

     763 .................TGGGTTCTGGATTGCGTATCAGAGAAAACCAA 794
                         :::|||   :::|||:::   :::|||   ||
     664 LeuArgGluLeuGluArgTyrValThrSerCysLeuArgLysLysArgLy 680

     795 AAAGGACGAAGCATTTTATCATCCTCGTCTGATTATTTATGAAGAGGAAG 844
         |
     680 s................................................... 680

     845 AAGAGGAAGAGGAAGAGGAGGAGGAAGTAGAAAATGAGGAACAAGATTCT 894
                     :::   |||:::|||:::            :::|||
     681 ...........ProGlnAlaGluLysValAspValIleAlaGlySerSer 693

     895 CAGAATGAGGGT......AGTACAGATGAGAAGTCAAGCCCTGCCAGCTC 938
         :::   :::|||      |||:::::::::::::||||||      |||||
     694 LysMetLysGlyPheSerSerSerGluSerGluSerSerSerGluSerSe 710

     939 TCAAGAAGGGTCTCCTTCAGGTGACCAGCAGTTTTCACCTAAATCCAACA 988
         |:::::   |||   :::::::::   :::::::::|||||||||
     710 rSerSerAspSerGluAspSerGluThrGluMetAlaProLysSer.... 725

     989 CTGAAAAATCTAAAGGGGAATTAATGTTTGATGATTCTTCAGATTCATCA 1038
              |||:::||||||:::
     726 .....LysLysLysGlyHis.............................. 730

    1039 CCGGAAAAACAGGAGAGA.....................AATTTAAA 1064
         |||   ::::::::::::                      ::::::::
     731 ProGlyArgGluGlnLysLysHisHisHisHisHisHisGlnGlnMetGl 747

    1065 CTGGACCCCGGCCGAAGTCCCACAGTTAGCTGCAGCAAAACGCAGGCTC: 1114
         :  :::|||||||  |||||||||                  :::   |
     747 nGlnAlaProAlaProValProGlnGlnProProProProProGlnGlnP 764

    1115 CT.....................CAGGGAAAGGAGCCTGGGTTGATTAAC 1143
         ||                     |||   :::::::|||
     764 roProProProProProProGlnGlnGlnGlnGlnPro............ 776

    1144 TTGTGTGCCAATGTCCCACCCGTCCCAGGTAACATTTTGCCCCCTGAGGT 1193
                        ||||||   |||         :::|||   :::
     777 ............ProProProProProProSerMetProGlnGln.. 788

    1194 CCGGGGTAA TTAATGGCTGCTGGACAAAACCTCCAAAGTTCTGAAAGAT 1243
                           |||:::      ::::::||||||
     789 ...................AlaAlaProAlaMetLysSerSerPro...P 798

    1244 CAGAAATGATAGCTACCTGGAGTCCAGCTGTACGGACACTGAGGAATATT 1293
              :::||||||||||     |||:::::::::
     798 roProPheIleAlaThrGlnValProValLeuGluProGlnLeuProGly 814

    1294 ACTAATAATGCTGACATTCAGCAGATGAACCGGCCATCAAATGTAGCACA 1343
         :::           |||   :::::::::::::|||      :::   ||
     815 SerValPheAspProIleGlyHisPheThrGlnPro......IleLeuHi 829

    1344 TATCTTACAGACTCTTTCAGCACCTACGAAAAATTTAGAACAGCAGGTGA 1393
         |:::   |||            |||      :::|||   |||      :
     829 sLeuProGlnProGluLeuProPro......HisLeuProGlnProProG 844

    1394 ATCACAGCCAGCAGGGACATACAAATGCCAATGCAGTGCTGTTTAGCCAA 1443
         ::||||||         |||   |||   :::|||||||:::
     844 luHisSerThrProProHisLeuAsnGlnHisAlaValValSerProPro 860

    1444 GTG..................................................... 1446
         :::
     861 AlaLeuHisAsnAlaLeuProGlnGlnProSerArgProSerAsnArgAl 877

    1447 .................................AAAGTGACTCCA.... 1458
                                         |||:::|||
     877 aAlaAlaLeuProProLysProAlaArgProProAlaValSerProAlaL 894

    1459 .....GAGACACACATGCTACAGCAGCAGCAGCAGGCCCAGCAGCAGCAG 1503
              :::|||   :::|||   |||      :::||||||      |||
     894 euThrGlnThrProLeuLeuProGlnProProMetAlaGlnProProGln 910

    1504 ...............CAGCAGCACCCGGTTTTACACCTTCAGCCCCAGCA 1538
                         ::::::   |||:::      |||      :::||
     911 ValLeuLeuGluAspGluGluProProAlaProProLeuThrSerMetGl 927

    1539 GATAATGCAGCTCCAGCAGCAGCAGCAGCAGCAGATC.............. 1575
         !   |||||||||      ||||||   |||::::::
     927 n...MetGlnLeuTyrLeuGlnGlnLeuGlnLysValGlnProProThrP 943
```

```
      1576 .......................TCTCAGCAACCTTACCCC...CAG 1596
                                  ||||||   |||   |||
       943 roLeuLeuProSerValLysValGlnSerGlnProProProProLeuPro 959

      1597 CAGCCGCCGCATCCATTTTCACAGCAGCAGCAGCAGCAACAGCAGCCACC 1646
              ||||||||||||      |||||||||   |||||||||   |||||
       960 ProProProHisProSerValGlnGlnGlnLeuGlnGlnGlnProProPr 976

      1647 ACCATCGCCT.......................CAGCAGCATCAG 1668
           ||||   |||                       ||||||||||||
       976 oProProProProGlnProGlnProProProGlnGlnGlnHisGln 991

seq_name: SwissProt_40:INVO_PIG

seq_documentation_block:
ID   INVO_PIG       STANDARD;      PRT;   347 AA.
AC   P18175;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   Involucrin.
GN   IVL.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90348475; PubMed=2385171;
RA   Tseng H., Green H.;
RT   "The involucrin genes of pig and dog: comparison of their segments of
RT   repeats with those of prosimians and higher primates.";
RL   Mol. Biol. Evol. 7:293-302(1990).
CC   -!- FUNCTION: INVOLUCRIN IS A KERATINOCYTE PROTEIN THAT FIRST APPEARS
CC       IN THE CELL CYTOSOL, BUT ULTIMATELY BECOMES CROSS-LINKED TO
CC       MEMBRANE PROTEINS BY TRANSGLUTAMINASE. ALL THAT RESULTS IN THE
CC       FORMATION OF AN INSOLUBLE ENVELOPE BENEATH THE PLASMA MEMBRANE.
CC   -!- TISSUE SPECIFICITY: PRESENT IN KERATINOCYTES OF EPIDERMIS AND
CC       OTHER STRATIFIED SQUAMOUS EPITHELIA.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M34441; AAA31058.1; -.
DR   InterPro; IPR002360; Involucrin.
DR   PROSITE; PS00795; INVOLUCRIN; 1.
KW   Keratinocyte; Repeat.
SQ   SEQUENCE   347 AA;  41408 MW;  3DE579871CF1C829 CRC64;


alignment_scores:
            Quality:  182.00              Length:    432
              Ratio:  0.958                Gaps:     16
 Percent Similarity:  43.981   Percent Identity: 22.454

alignment_block:
 US-09-664-641-10 x INVO_PIG   ..

 Align seg 1/1  to: INVO_PIG  from: 1  to: 347

      724 GCTTTAAAGCGAGCAAGTATTAAAATTGTGACTCCTGACTGGGTTCTGGA 773
          ||||||::::::      :::|||   |||:::|||          ||
       14 AlaLeuSerGlnGluProLeuLysProValSerProPro......AlaAs 28

      774 TTGCGTATCAGAGAAAACCAAAAAG..................... 798
          |      :::|||::::::|||:::
       28 pThrGlnGlnGluGlnValLysGlnAlaThrProLeuProAlaProCysG 45

      799 .....................................GACGAAGCATTT 810
                                              ::::::|||
       45 lnLysMetLeuSerGluLeuProValGluValProLeuGluAspAlaGlu 61

      811 TATCATCCTCGTCTGATT......ATTTATGAAGAGGAA.....GA 845
          |||      |||:::      :::   |||:::|||        ::
       62 LysHisThrThrLeuValLysGlyValProGluGlnGluCysGluProGl 78

      846 AGAGGAAGAGGAAGAGGAGGAGGAAGTAGAAAATGAGGAACAAGATTCTC 895
          :   :::|||   :::::::::|||::::::   |||:::|||:::::::|
       78 nProGlnGluProGlnGlnGlnGluLeuHisValGluGlnGlnGlnGlnG 95

      896 AGAATGACGGTAGT..............ACAGATGAGAAGTCAAGCCCT 930
          ||:::|||:::::                :::|||:::::::::::::
       95 lnGlnGluSerGlnValGlnGluLeuHisValAspGlnGlnGlnGlnGln 111

      931 GCCAGCTCTCAAGAAGGGTCTCCTTCAGGTGACCAGCAGTTTTCACCTAA 980
             :::|||||||||   :::         |||||||||          ::
      112 GlnGluSerGlnGluGlnGluLeuHisValAspGlnGln.....Gl 125

      981 ATCCAACACTGAAAAATCTAAAGGGGAATTAATGTTTGATGATTCTTCAG 1030
          :::::::   |||:::::::::   ||||||      |||:::::::::
      125 nGlnGlnGlnGluSerGlnGluGlnGluLeuHisValAspGlnGlnGlnG 142

     1031 ATTCATCACCGGAAAAACAGGAGAGAAATTTAAACTGGACCCCGGCCGAA 1080
          ::::::::   :::   ||||||
      142 lnGlnGluSerGlnValGlnGlu..................... 149

     1081 GTCCCACAGTTAGCTGCAGCAAAACGCAGGCTGCCTCAGGGAAAGGAGCC 1130

      149 ..................................................... 149

     1131 TGGGTTGATTAACTTGTGTGCCAATGTCCCACCCGTCCCAGGTAACATTT 1180

      149 ..................................................... 149

     1181 TGCCCCCTGAGGTCCGGGGTAATTTAATGGCTGCTGGACAAAACCTCCAA 1230
                                      :::   :::|||::::::   |||
      150 .......................LeuHisValGlyHisHisGlnGln 157

     1231 AGTTCTGAAAGATCAGAAATGATAGCTACCTGGAGTCCAGCTGTACGGAC 1280
          ::::::|||   :::|||
      158 GlnGlnGluSerGlnGlu..................... 163

     1281 ACTGAGGAATATTACTAATAATGCTGACATTCAGCAGATGAACCGGCCAT 1330
             :::::::::      :::       ||||||::::::      :
      164 ....GlnGluLeuHisValAspHisHisGlnGlnGlnGlnGluSerGlnG 179

     1331 CAAATGTAGCACATATCTTACAGACTCTTTCAGCACCTACGAAAAATTTA 1380
          :::::      |||:::   |||
      179 luGlnGluLeuHisValAspGln..... 186

     1381 GAACAGCAGGTGAATCACAGCCAGCAGGGACATACAAATGCCAATGCAGT 1430
          :::|||||||   ::::::|||||||:::
      187 GlnGlnGlnGlnGlnGluSerGlnGlu..... 195

     1431 GCTGTTTAGCCAAGTGAAAGTGACTCCAGAGACACACATGCTACAGCAGC 1480
                                    |||   |||:::   |||||||
      196 .......................GlnGluLeuHisValAspGlnGlnG 204

     1481 AGCAGCAGGCCCAGCAGCAGCAGCAGCAGCACCCGGTTTTACACCTTCAG 1530
          ||||||||   :::::::|||:::|||:::      |||   |||   |||
      204 lnGlnGlnGlnGluSerGlnGluGlnGluLeuHisValAspHisHisGln 220

     1531 CCCCAGCAGATAATGCAGCTCCAGCAG..............CAGCAGCA 1565
             ||||||      |||:::|||:::                ||||||||
      221 GlnGlnGlnGluSerGlnValGlnGluLeuHisValAspHisGlnGlnGl 237

     1566 GCAGCAGATCTCT.....................................C 1579
          |||||||   |||                                    |
      237 nGlnGlnGlnSerGlnGluGlnGluLeuHisValAspGlnHisGlnGlnG 254

     1580 AGCAACCTTACCCCCAGCAGCCGCCGCATCCATTTTCACAGCAGCAG... 1626
```

```
      |||||         :::|||      |||      :::||||||||||
  254 lnGlnGluSerGlnGluGlnGluLeuHisValAspGlnGlnGlnGlnGlu 270

 1627 ..................CAGCAGCAACACCAGCCACCACCATCGCCTCA 1658
                        ||||||||||||||||              ||
  271 LeuGlnValGlnGluValGlnGlnGlnGlnGln.............GlnGl 283

 1659 GCAGCATCAGCTTTTTGGACATGATCCAGCAGTGGAGATTCCAGAAGAAG 1708
      ||||:::|||          |||:::   |||   ::::::   :::|||
  283 nGlnGluGlnGlnGluAspHisGlnLysAlaGluHisLeuGluGlnGlu. 299

 1709 GCTTCTTATTGGGATGTGTGTTTGCAATTGCGGATTATCCAGAGCAGATG 1758
                                          |||   :::
  300 .......................................GluAlaGln 302

 1759 TCTGATAAGCAACTGCT.GGCCACCTGGAAAAGGATAATCCAGGCACATG 1807
         ::::::||||||   |||:::||||||::::::::   |||:::::
  303 ArgGluGlnGlnLeuLysGlyGlnLeuGluGlnGluLysLysGlyValTy 319

 1808 GCGGCACTGTTGACCCCACCTTCACGAGTCGATGCACGCACCTT 1851
      :::|||| :::|||      :::|||:::|||      ||||||
  319 rGlnHis.LeuAspGlnGluLeuThrLysArgAspGluHisLeu 333
```

seq_name: SwissProt_40:KAPC_DICDI

seq_documentation_block:

```
ID   KAPC_DICDI     STANDARD;      PRT;   648 AA.
AC   P34099;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   cAMP-dependent protein kinase catalytic subunit (EC 2.7.1.37).
GN   PKAC OR PK2 OR PK3.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91323730; PubMed=1864510;
RA   Buerki E., Anjard C., Scholder J.-C., Reymond C.D.;
RT   "Isolation of two genes encoding putative protein kinases regulated
RT   during Dictyostelium discoideum development.";
RL   Gene 102:57-65(1991).
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE=93385090; PubMed=8373760;
RA   Anjard C., Etchebehere L., Pinaud S., Veron M., Reymond C.D.;
RT   "An unusual catalytic subunit for the cAMP-dependent protein kinase
RT   of Dictyostelium discoideum.";
RL   Biochemistry 32:9532-9538(1993).
RN   [3]
RP   CHARACTERIZATION.
RC   STRAIN=AX3;
RX   MEDLINE=93066311; PubMed=1332055;
RA   Mann S.K.O., Yonemoto W.M., Taylor S.S., Firtel R.A.;
RT   "DdPK3, which plays essential roles during Dictyostelium development,
RT   encodes the catalytic subunit of cAMP-dependent protein kinase.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:10701-10705(1992).
CC   -!- FUNCTION: ESSENTIAL FOR DIFFERENTIATION AND FRUIT MORPHOGENESIS.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC   -!- SUBUNIT: IN DICTYOSTELIUM THE HOLOENZYME IS A DIMER COMPOSED OF
CC       A REGULATORY (R) AND A CATALYTIC (C) SUBUNIT. IN THE PRESENCE OF
CC       CAMP IT DISSOCIATES INTO THE ACTIVE C SUBUNIT AND AN R MONOMER.
CC   -!- DEVELOPMENTAL STAGE: CAPK ACTIVITY IS LOW IN VEGETATIVELY GROWING
CC       AMOEBAE, INCREASES DURING DEVELOPMENT OF AGGREGATION AND REACHES
CC       A MAXIMUM AT CULMINATION.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC       CAMP SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M38703; -; NOT_ANNOTATED_CDS.
DR   PIR; JQ1150; JQ1150.
DR   HSSP; P05132; 1ATP.
DR   DictyDb; DD02030; pkaC.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR000961; Pkinase_C.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00433; pkinase_C; 1.
DR   SMART; SM00133; S_TK_X; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding; cAMP;
KW   Phosphorylation.
FT   DOMAIN       58     64       ASN-RICH.
FT   DOMAIN      136    223       GLN-RICH.
FT   DOMAIN      233    250       THR-RICH.
FT   DOMAIN      336    590       PROTEIN KINASE.
FT   NP_BIND     342    350       ATP (BY SIMILARITY).
FT   BINDING     365    365       ATP (BY SIMILARITY).
FT   ACT_SITE    459    459       BY SIMILARITY.
FT   MOD_RES     490    490       PHOSPHORYLATION (BY SIMILARITY).
SQ   SEQUENCE   648 AA;  74458 MW;  D0F9B3A48C58D084 CRC64;
```

alignment_scores:

```
           Quality:  181.50              Length:    626
             Ratio:  0.726                Gaps:     21
 Percent Similarity:  39.936   Percent Identity: 21.246
```

alignment_block:

US-09-664-641-10 x KAPC_DICDI ..

Align seg 1/1 to: KAPC_DICDI from: 1 to: 648

```
 1027 TCAGATTCATCACCGGAAAAACAGGAGAGAAATTTAAACTGGACCCCGGC 1076
      |||:::|||:::   ::::::::::::   |||   |||   |||
    2 SerAsnSerAsnAsnAsnSerSerSerGlyAsnHisAsnSerThrThrIl 18

 1077 CGAAGTCCCACAGTTAGCTGCAGCAAAACGCAGGCTGCCTCAGGGAAAGG 1126
       :::   |||:::
   18 eAsnAsnProLys..................................... 22

 1127 AGCCTGGGTTGATTAACTTGTGTGCCAATGTCCCACCCGTCCCA..... 1170
                 :::|||:::   :::|||:::|||
   23 ..........ValAsnValTyrSerAsnIleProAsnSerThrThrTyr 35

 1171 .................GGTAACATTTTGCCCCCTGAGGTCCGGGGTAA 1202
                        |||:::   |||      :::      :::||
   36 ThrTyrGlySerGlyGlyGlyGlyThrLeuSerGlyAsnAsnThrAsnAs 52

 1203 TTTAATGGCTGCTGGACAAAACCTCCAAAGTTCTGAAAGATCAGAAATGA 1252
      |      :::   ::::::|||   ::::::::::::   |||
   52 nAsnAsnThrAsnAsnAsnAsnAsnAsnAsnAsnSerSerGlyAspA 69

 1253 TAGCTACCTGGAGTCCAGCTGTACGGACACTGAGGAATATTACTAATAAT 1302
       :::   ||||||::::::   :::      ::::::|||   :::
   69 snLysAsnHisSerProValThrSerAlaThrAspArgLeuThr...Lys 84

 1303 GCTGACATTCAGCAGATG..............................AA 1322
      ||||||::::::                                    ::
   85 MetAspIleGluGluLysTrpAspAsnLysAsnTyrGluLysAspGluAr 101

 1323 CCGGCCATCAAATGTAGCACATATCTTACAG.................... 1353
      ::::   |||   :::   ||||||||
  101 gGluLysSerProLeuPheHisIleLeuAlaSerAsnLeuAsnSerPheG 118
```

```
1354 ..........................ACTCTTTCAGCACCTACGAAAAAT 1377
                               :::|||:::   |||      |||
 118 lyAsnPheLysValProSerThrPheSerLeuThrProProGluProAsn 134

1378 TTAGAACAGCAGGTGAATCACAGCCAGCAGGGACATACAAATGCCAATGC 1427
     :::||||||   :::::::::   |||   :::
 135 LysGlnGlnGlnProGlnGlnGlnProGlnGlnGln.............. 146

1428 AGTGCTGTTTAGCCAAGTGAAAGTGACTCCAGAGACACACATGCTACAGC 1477
             :::      :::      |||:::   ::::::   ||||
 147 ..........GlnProGlnGlnGlnGlnProGlnGlnGlnGlnProGlnG 160

1478 AGCAGCAGCAGGCCCAGCAGCAGCAGCAGCAGCACCCGGTTTTACACCTT 1527
     ||||||||||||   |||   ||||||||||||:::|||      :::|||
 160 lnGlnGlnGlnGlnGlnProGlnGlnGlnGlnGlnProGlnGlnGlnLeu 176

1528 CAGCCCCAGCAGATAATGCAGCTCCAGCAGCAGCAGCAGCAGCAGATCTC 1577
     |||   ::::::   :::|||   ||||||   ||||||||||||:::::
 177 GlnGlnAsnAsnGlnGlnGlnGlnGlnGlnLeuGlnGlnGlnGlnLeuGl 193

1578 TCAGCAACCTTACCCCCAGCAGCCGCCGCATCCATTTTCACAGCAGCAGC 1627
     :||||||          ||||||      :::      :::||||||||||
 193 nGlnGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnG 210

1628 AGCAGCAACAGCAGCCACCACCATCGCCTCAGCAGCATCAGCTTTTTGGA 1677
     ||||||||:::|||      :::   ||||||:::|||   :::
 210 lnGlnGlnLysGlnGlnLysGlnGlnGlnGlnGlnGlnGlnHisLeuHis 226

1678 CATGATCCAGCAGTGGAGATTCCAGAAGAAGGCTTCTTATTGGGATGTGT 1727
     :::|||      |||:::   |||
 227 GlnAspGlyIleValAsnThrPro..................... 234

1728 GTTTGCAATTGCGGATTATCCAGAGCAGATGTCTGATAAGCAACTGCTGG 1777
                                    |||      |||      :
 235 ............................SerThrThrGlnThrSerT 241

1778 CCACCTGGAAAAGGATAATCCAGCCACATGGCGGCACTGTTGACCCCACC 1827
     ::|||                          |||:::      |||
 241 hrThr.........................ThrThrThrThrThrThr 248

1828 TTCACGAGTCGATGCACGCACCTTCTCTGTGAGAGTCAAGTCAGCAGCGC 1877
        |||:::      |||      |||      :::::::::   ||||::
 249 ThrThrAsnProHisThrSerGlyLeuSerLeuGlnHisAlaHisSerSe 265

1878 GTATGCACAGGCAATAAGAGAAAGAAAGAGATGTGTTACTGCACACTGGT 1927
     :|||:::
 265 rTyrThr..................................... 267

1928 TAAACACAGTCTTAAAAAAAAAAAAAAATGGTACCGCCGCACCGAGCCCTT 1977
                                       |||   ::::::|||
 268 ................................ProSerAsnValLeu 272

1978 CACTTCCCAGTGGCCTTCCCACCAGGAGGAAAGCCATGTTCACAGCATAT 2027
     |||   |||:::   |||      ::::::   |||
 273 HisSerProThrHisPhe...GlnSerSerLeuProThrArgLeuAspTh 288

2028 TATTTCTGTGACTGGATTTGTTGATAGTGACAGAGATGACCTAAAATTAA 2077
            :::|||      :::   :::::::::
 288 rAsnProIleThrThrProIleArgGlnGlnGln................ 299

2078 TGGCTTATTTGGCAGGTGCCAAATATACGGGTTATCTATGCCGCAGCAAC 2127

 299 ..................................................... 299

2128 ACAGTCCTCATCTGTAAAGAACCAACTGGTTTAAAGTATGAAAAAGCCAA 2177
                    ::::::         |||:::   ::::::   ::
 300 ..............GlnSerGlnGlnGlnLeuGlnGlnGlnGlnLeuGl 311

2178 AGAGTGGAGGATACCCTGTGTCAACGCCCAGTGGCTTGGCGACATTCTT. 2226
     ::::         |||   |||||||:::   :::|||      :::
 311 nGlnIleProProProThrValAsnSerPhePheLeuProProProValA 328

2227 .............CTGGGAAACTTTGAGGCACTGAGGCAGATTCAGTAT 2262
                          |||   :::|||:::   :::|||   :::
 328 snAlaArgGluArgLeuLysGluPheLysGlnIleArgValLeuGlyThr 344

2263 AGTCGCTACACGGCATTCAGTCTGCAGGATCCATTTGCCCCTACCCAGCA 2312
     :::   :::                                      ::
 345 GlyThrPheGlyLysVal.............................Ty 351

2313 TTTAGTTTTAAATCTTTTAGATGCTTGGAGAGTTCCCTTAAAAGTGTCTG 2362
     :|||:::   |||      |||:::             :::|||
 351 rLeuIleGlnAsnThrLysAspGlyCysTyrTyrAlaMetLys....... 365

2363 CAGAGTTGTTGATGAGTATAAGACTACCTCCCAAACTGAAACAGAATGAA 2412
             ::::::                  :::||||||||||   |||
 366 ........CysLeuAsnLysAlaTyrValValGlnLeuLysGlnValGlu 379

2413 GTAGCTAATGTCCAGCCTTCTTCCCAAAGAGCCAGAATTGAAGACGTACC 2462
          |||   :::   |||
 380 HisLeuAsnSerGluLysSerIle........................... 387

2463 ACCTCCCACTAAAAAGCTAACTCCAGAATTGACCCCTTTTGTG......C 2506
                    |||:::          |||||||:::   |
 388 ................LeuSerSerIleHisHisProPheIleValAsnL 399

2507 TTTTCACTGGATTCGAGCCTGTCCAGGTTCAACAGTATATTAAGAAGCTC 2556
     ||:::   :::|||                        |||      |||||||||
 399 euTyrGlnAlaPhe.................GlnAspGluLysLysLeu 409

2557 TACATTCTT..............GGTGGAGAGGTTGCGGAGTCTGCACA 2591
          |||:::|||          ||||||||||
 410 TyrLeuLeuPheGluTyrValAlaGlyGlyGluVal.............. 421

2592 GAAGTGCACACACCTCATTGCCAGCAAA......GTGACTCGCACCGTGA 2635
          |||||||||      |||              :::   |||:::|
 422 ....PheThrHisLeuArgLysSerMetLysPheSerAsnSerThrAlaL 437

2636 AGTTCCTGACGGCGATTTCTGTCGTGAAGCACATAGTGACGCCAGAGTGG 2685
     |||||   :::|||
 437 ysPheTyrAlaAla..................................... 441

2686 CTGGAAGAATGCTTCAGGTGTCAGAAGTTCATTGATGAGCAGAACTACAT 2735
          |||          :::   :::|||:::   :::||||||   ::
 442 ......GluIleValLeuAlaLeuGluPheLeuHisLysGlnAsnIleVa 456

2736 TCTCCGAGATGCTGAGGCAGAAGTACTT 2763
     :   ||||||   :::   |||   |||
 456 lTyrArgAspLeuLysProGluAsnLeu 465
```

seq_name: SwissProt_40:GALY_KLULA

seq_documentation_block:

ID GALY_KLULA STANDARD; PRT; 1008 AA.
AC P32257;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Transcription regulatory protein GAL11.
GN GAL11.
OS Kluyveromyces lactis (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID=28985;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92020226; PubMed=1923818;
RA Dickson R.C., Hopper J., Mylin L.M., Gerardot C.J.;
RT "Sequence conservation in the Saccharomyces and Kluveromyces GAL11
RT transcription activators suggests functional domains.";
RL Nucleic Acids Res. 19:5345-5350(1991).
CC -!- FUNCTION: AUXILIARY TRANSCRIPTION ACTIVATOR FOR GENES ENCODING
CC GALACTOSE-METABOLIZING ENZYMES. ESSENTIAL FOR NORMAL GROWTH ON
CC NONFERMENTABLE CARBON SOURCES, FOR SPORULATION AND MATING.
CC -!- MISCELLANEOUS: GAL11 LACKS A DNA-DOMAIN, IT PROBABLY COMPLEXES
CC WITH GAL4 THAT HAS THE CAPACITY TO BIND DNA. ASSOCIATION BETWEEN

```
CC       GAL11 AND GAL4 MAY SERVE TO EXPEDITE PHOSPHORYLATION OF GAL4.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M68870; AAA35254.1; -.
KW   Transcription regulation; Activator; Galactose metabolism.
FT   DOMAIN       75     84       POLY-GLN.
FT   DOMAIN      257    287       GLN-RICH.
FT   DOMAIN      364    411       GLN-RICH.
FT   DOMAIN      456    464       POLY-GLN.
SQ   SEQUENCE   1008 AA;  114831 MW;  436D1EBAEA17F7DB CRC64;


alignment_scores:
            Quality:  180.50              Length:    775
              Ratio:  0.520                Gaps:     39
 Percent Similarity:  44.774   Percent Identity: 21.032

alignment_block:
 US-09-664-641-10 x GALY_KLULA   ..

 Align seg 1/1  to: GALY_KLULA  from: 1  to: 1008

     693 TCCAGAGCCAAAGGGGGGGAA..............................  713
         ||||||   :::||||||||||
       2 SerArgIleAsnGlyGlyGluLysAlaAsnPheAsnLeuGluLysLeuLy   18

     714 ....ATACGAATGTGCTTTAAAGCGAGCAAGTATTAAAATTGTGACTCC.  758
                    |||:::::::|||   |||   :::|||:::|||
      18 sLysSerAlaGluGlnPheGluSerSerLeuTyrAlaSerCysAsnSerL   35

     759 ................TGACTGGGTTCTGGATTGCGTATCAGAGAAAAC  791
                           :::   :::   :::          ::::::
      35 ysAspLeuTyrLeuAspAlaMetArgLysArgIleAlaAlaMetAspSer   51

     792 CAAAAAGGACGAAGCATTTTATCATCCTCGTCTGATTATTTATGAAGAGG  841
           |||   :::||||||   :::::: |||            ::::::
      52 AlaLysArgLysSerIleGluAlaGln.Arg.............GlnLysL   64

     842 AAGAAGAGGAAGAGGAAGACGAGGAGGAAGTAGAAAATCAGGAACAAGAT  891
           :::   :::   :::   :::   |||:::::::::::::|||:::
      64 euAlaGlnAlaGlnAlaGlnAlaGlnAlaValGlnArgGlnGlnGlnGln   80

     892 TCTCAGAATGAGGGTAGTACAGATGAGAAGTCAAGCCCTGCCAGCTCTCA  941
         :::|||::::::         :::::::::::|||      ::::::
      81 GlnGlnGlnGlnHisHisMetGlnGlnGlnGlySerGlyGlnGlnGlnAl   97

     942 AGAAGGGTCTCCTTCAGGTGACCAGCAGTTTTCACCTAAATCCAACACTG  991
           :::|||   |||:::                  :::   |||::::
      97 aAsnGlyIleProAla.....................AsnIleAsnAlaG  107

     992 AAAAATCTAAAGGGGAATTAATGTTTGATGATTCTTCAGATTCATCACCG 1041
           ::                  ||||||   :::::::::   ::::::
     107 ln..................MetPheLeuAsnGlnGlnAlaGlnAlaArg  117

    1042 GAAAAACAGGAGAGAAATTTAAACTGGACCCCGGCCGAAGTCCCACAGTT 1091
         ::::::   :::|||:::::::::   |||                ||
     118 GlnGlnAlaGlnArgGlnIleArgSerThr...................Le  128

    1092 AGCTGCAGCAAAACGCAGGCTGCCTCAGGGAAAGGAGCCTGGGTTGATTA 1141
         |::::::   :::::::::   |||:::   ::::::|||
     128 uThrGlyGlnGlnGlnGlnGlnProMetProGlnGlnPro.......... 141

    1142 ACTTGTGTGCCAATGTCCCACCCGTCCCAGGTAACATTTTGCCCCCTGAG 1191
                                                 |||   :::
     142 .....................................GlnProGlnGln  145

    1192 GTCCGGGGTAATTTAATGGCTGCTGGACAAAACCTCCAAAGTTCT..... 1236
           :::   |||:::|||           :::||||||::::::
     146 GlnGlnProAsnMetMetArgProGlnLeuThrLeuGlnGlnGlnGlnGl  162

    1237 .....................................GAAAGATCAGAAA 1249
                                                   |||   :::
     162 nLeuAlaAsnGluLeuLysValThrGluIleProArgGluLeuLeuGlnL  179

    1250 TGATAGCTACCTGGAGTCCAGCTGTACGGACACTGAGGAATATTACTAAT 1299
           |||   :::     |||::::::   |||     :::||||||
     179 ysIleProAsnLeuProProGlyIleThrThrTrpProGlnIleThrPhe  195

    1300 AATGCTGACATTCAGCAGATGAACCGGCCATCAAATGTAGCACATATCTT 1349
           |||         |||  ||||||||||||:::         |||::
     196 TrpAla.........GlnLysAsnArgProSerGlnAsnAspLeuIleIl  209

    1350 ACAGACTCTTTCAGCACCTACGAAAAATTTA............GAACAGC 1387
         :                |||||::::::                :::|||
     209 e.................ThrLysLysValTyrGlnLeuHisGlnGlnL  220

    1388 AGGTGAATCACAGCCAGCAGGGACATACAAATGCCAATGCAGTGCTGTTT 1437
           :::|||   |||:::      ::::::|||:::|||   :::
     220 euLeuAsnLysSerLysLeuGlnGlnSerAsnValAsnArgMet...... 234

    1438 AGCCAAGTGAAAGTGACTCCAGAGACACACATGCTACAGCAGCAGCAGCA 1487
                              :::|||   :::   |||      :::
     235 .....................AsnThrSerValGlnGlnProGlyMetGl  244

    1488 GGCCCAGCAGCAGCAGCAGCAGCACCCGGTTTTACACCTTCAGCCCCAGC 1537
          :::|||||:::   :::|||:::      ::::::      |||||||
     244 yValGlnGlnSerProAsnGlnArgMetAsnMetAsnGlnValProGlnG  261

    1538 AGATAATGCAGCTCCAGCAGCAGCAGCAGCAGCAGATCTCTCAGCAACCT 1587
         ||   :::|||:::::::|||||||||||||||||||   :::|||
     261 lnGlnGlnGlnValArgGlnGlnGlnGlnGlnGlnGlnGlnGln...... 275

    1588 TACCCCCAGCAGCCGCCGCATCCATTTTCACAGCAGCAGCAGCAGCAACA 1637
                                  ||||||||||||:::::||
     276 ..........................GlnGlnGlnGlnHisLysGl  282

    1638 GCAGCCACCACCATCGCCTCAGCAGCATCAGCTTTTTGGACATGATCCAG 1687
         |||||||              |||||||||          :::   ||||
     282 nGlnPro..............GlnHisGln.........GluProProA  291

    1688 CAGTGGAGATTCCAGAAGAAGGCTTCTTATTGGGATGTGTGTTTGCAATT 1737
         |||||                        ||||||   :::       :::
     291 laVal.......................LeuGlyArgLeuAsnGlnVal  299

    1738 GCGGATTATCCAGAGCAGATGTCTGATAAGCAACTGCTGGCCACCTGGAA 1787
                       ||||||          |||:::|||        :::   ||
     300 PheThrGlnAlaGluGln.........LysMetLeuTyrGluAlaGlyLy  313

    1788 AAGGATAATCCAG......GCACATGGCGGCACTGTTGACCCCACCTTCA 1831
         |::::::|||:::          ::::::|||   :::   |||::::::
     313 sLysMetIleGluAsnLeuGlnArgSerGlyLysLeuProProAsnLeu.  329

    1832 CGAGTCGATGCACGCACCTTCTCTGTGAGAGTCAAGTCAGCAGCGCGTAT 1881
                                  :::||||||   :::      |||
     330 ..........................AspSerGlnGlnGlnLeuLeuTyr  337

    1882 GCACAGGCAATAAGAGAAAGAAAGAGATGTGTTACTGCACACTGGTTAAA 1931
           :::                                    ::::::||
     338 IleLys..................................LysTyrMetAs  343

    1932 CACAGTCTTAAAAAAAAAAAAAATGGTACCGCCGCACCGAGCCCTTCACT 1981
         |   ::::::   ||||||                :::   |||   :
     343 nGlnMetValLeuLysLysMet..............GlnGlnLeuGlyI  355

    1982 TCCCAGTGGCCTTCCCACCAGGAGGAAAGCCATGTTCACAGCATATTATT 2031
         ::|||:::   :::|||   :::::::::      ||||||   :::
     355 leProIleProMetProGlySerAlaGlnGlnGlnSerGlnLeuLeuGln  371

    2032 TCTGTGACTGGATTTGTTGATAGTGACAGAGATGACCTAAAATTAATGGC 2081
```

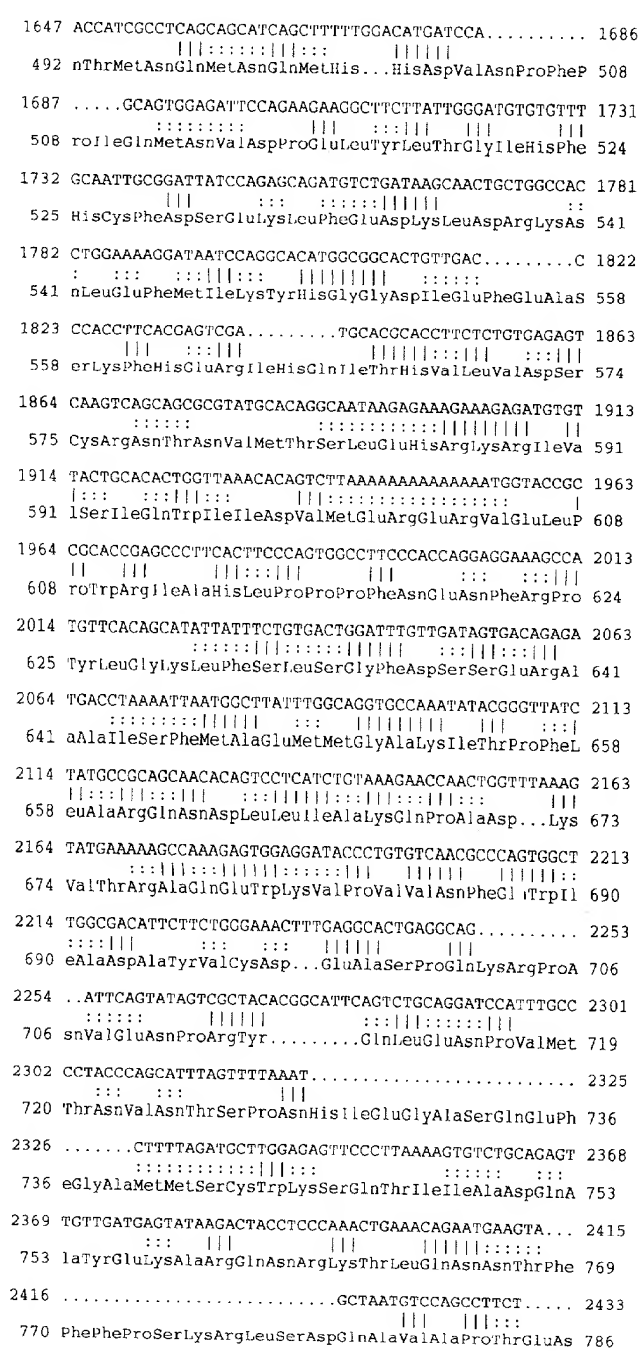
```
 372 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnArgGlnGlnGl 388

2082 TTATTTGGCAGGTGCCAAATATACGGGTTATCTATGCCGCAGCAACACAG 2131
 388 n.IleProGlnVal.GlnGlnAlaProLeuGlnGlnProGlnValGlnPr 404

2132 TCCTCATCTGTAAAGAACCAACTGGTTTAAAGTATGAAAAAGCCAAAGAG 2181
 404 oPro.....GlnSerGlnGlnA 410

2182 TCGAGGATACCCTGTGTCAACGCCCAGTCCTTGGCGACATTCTTCTGGG 2231
 410 laGln.....SerArgArg 414

2232 AAACTTTGAGGCACTGAGGCAGATTCAGTATAGTCGCTACACGGCATTCA 2281
 415 GlnAlaMetGlyMetLysProThrProVal.....IlePr 426

2282 GTCTGCAGGATCCATTTGCCCCTACCCAGCATTTAGTTTTAAATCTTTTA 2331
 426 oAsnAlaGlyValMetLysProGlnProGln..... 436

2332 GATGCTTGGAGAGTTCCCTTAAAAGTGTCTGCAGAGTTGTTGATGAGTAT 2381
 437 .....GlnThr.ProGlnAsnIleGlnGlnProMetMetGlnGlnGl 450

2382 AAGACTACCTCCCAAACTGAAACAGAATGAAGTAGCTAATGTCCAGCCTT 2431
 450 nSerSerProProProGlnGlnGln.....GlnGlnG 461

2432 CTTCCCAAAGAGCCAGAATTGAAGACGTACCACCTCCCACT..... 2472
 461 lnGlnGlnGlnSerLysIle...SerValProArgProThrGluGlnAsp 476

2473 .....AAAAAGCTAACTCCAGAATTGACC......CCTTTTGT 2504
 477 MetLeuAlaLeuLysLysLeuAsnAlaGluValThrLysAsnProValLy 493

2505 GCTT.....TTCACTGGATTCGAGCCTGTCCAG.... 2532
 493 sLeuAsnAspIleThrAsnArgLeuThrAsnAspGlnLysGlnGlnIleA 510

2533 .....GTTCAACAGTAT 2544
 510 rqSerLysLeuHisAlaAsnGlnGlnLeuPheSerSerValGluSerPhe 526

2545 ATTAAGAAGCTCTACATTCTTGGTGGAGAGGTTGCGGAGTCTGCACAG.. 2592
 527 IleProThrLeuTyrMetLeuThrArgAsnGluGluHisIleArgGlnLe 543

2593 .....AAGTGCACACACCTCATTGCCAGCAAAGTGACTCGCA 2629
 543 uLeuGlnIleArgMetLeuThrArgGluIleMetGluLysAlaValArgG 560

2630 CCGTGAAGTTCCTGACGGCGATTTCTGTCGTGAAGCACATAGTGACGCCA 2679
 560 lyVal...PheLeuValGluProAsnValValAspLysValIle..... 573

2680 GAGTGGCTGGAAGAATGCTTCAGGTGTCAGAAGTTCATTGATGAGCAGAA 2729
 574 .....PheArgTyrGlnLysTyr..... 579

2730 CTACATTCTCCGAGATGCTGAGGCAGAAGTACTTTTCTCTTTCAGCTTGG 2779
 580 .....TyrGluTyrThrLysG 585

2780 AAGAATCCTTAAAACGG 2796
 585 luGlnLeuLeuArgArg 590
```

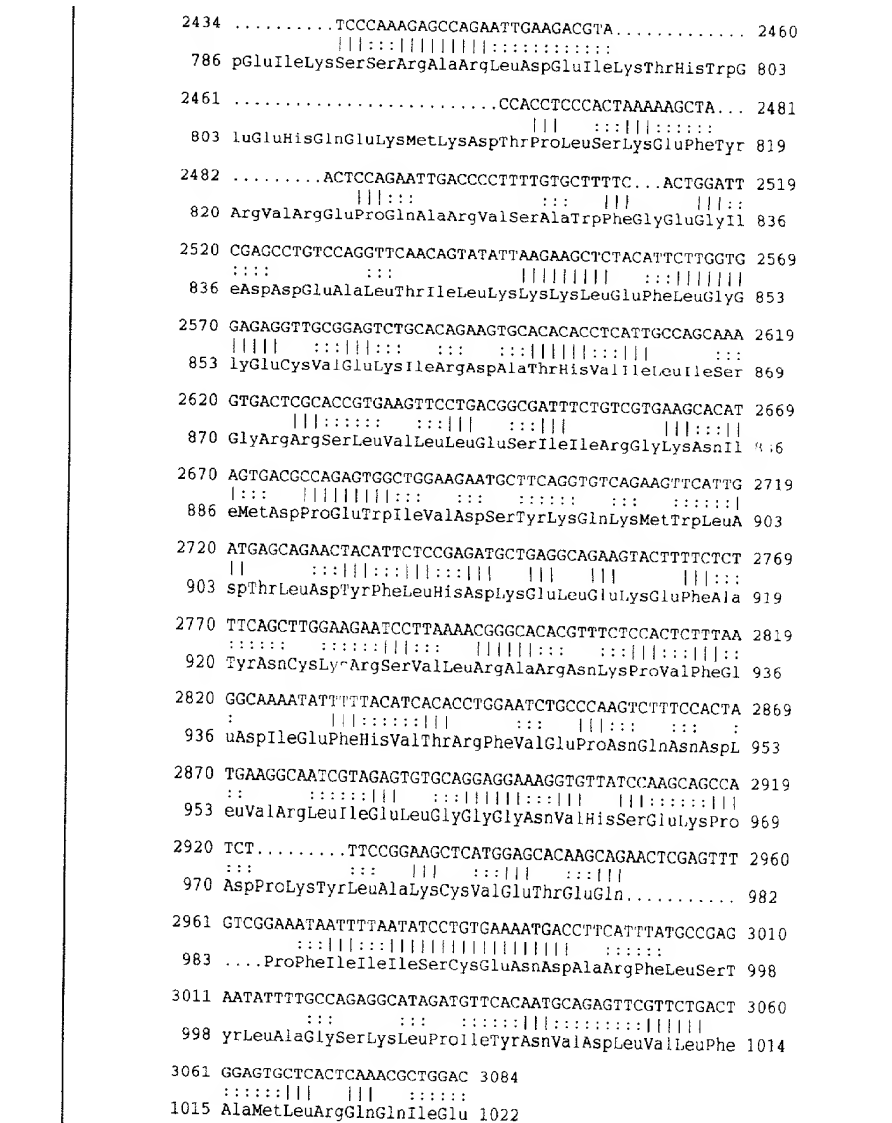
```
OM of: US-09-664-641-10 to: PIR_71:*   out_format : pfs

Date: Jun 4, 2002  5:36 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_n2p.model -DEV=x1p
 -Q=/cgn2_1/USPTO_spool/US09664641/runat_04062002_110932_9184/app_query.fasta_1.3697
 -DB=PIR_71 -QFMT=fastan -SUFFIX=rpr -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500
 -MINLEN=0 -MAXLEN=2000000000 -USER=US09664641_@CGN1_1_173
 -NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-664-641-10
Query length: 3580
Database: PIR_71:*
Database sequences: 283138
Database length: 96089334
Search time (sec): 144.160000
```

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
pir2:T24877	+	524.00	638.78	1.6e-28	1074	! hypothetical protein T13F2.3a
pir2:T24887	+	524.00	638.76	1.6e-28	1076	! hypothetical protein T13F2.3b
pir2:T37978	+	314.50	377.81	6.7e-14	878	! multiple BRCT domain containing
pir2:G85039	+	265.50	317.60	1.7e-10	765	! hypothetical protein AT4g03130
pir2:T49721	+	229.00	271.05	6.3e-08	831	! hypothetical protein B23L21.34C
pir1:RGBYS5	+	223.50	263.37	1.5e-07	905	! regulatory protein SNF5 - yeast
pir2:S69205	+	220.50	257.16	2.6e-07	1180	! stripe a/b protein - fruit fly
pir2:T14577	+	219.50	253.96	3.2e-07	1457	! protein kinase YakA (EC 2.7.1.
pir2:S37876	+	213.00	252.07	8.1e-07	738	! glutamine-rich protein YKL054c
pir2:T13167	+	210.50	246.05	1.3e-06	1010	! Lola-like protein - fruit fly
pir2:S44898	+	204.50	238.61	3.3e-06	1000	! ZK1236.3 protein - Caenorhabdi
pir2:A38346	-	200.00	246.50	5.3e-06	230	! ultra-high-sulfur keratin 1 - m
pir2:B38346	-	198.00	244.27	7.3e-06	223	! ultra-high-sulfur keratin 2 - m
pir2:T14004	+	196.50	225.54	1.3e-05	1390	! trfA protein - slime mold (Dic
pir2:T03455	+	196.50	213.84	1.6e-05	4957	! ALR protein - human
pir2:T03454	+	196.50	213.29	1.6e-05	5262	! ALR protein - human
pir2:T13828	+	195.50	216.64	1.8e-05	3190	! CREB-binding protein homolog +
pir2:T15651	-	192.50	238.94	1.7e-05	188	! hypothetical protein C27A2.5 -
pir2:A34599	+	192.00	234.45	2.0e-05	286	! DNA-binding protein MCM1 - yeas
pir2:T24272	-	191.50	238.95	2.0e-05	164	! hypothetical protein T01B7.8 -
pir2:S60966	+	189.50	222.37	3.6e-05	756	! probable protein kinase YNL161w
pir2:T30608	+	188.50	217.07	4.5e-05	1173	! proteophosphoglycan homolog 6I
pir2:T13052	+	188.50	215.55	4.7e-05	1383	! guanine nucleotide exchange fa
pir2:B53225	+	188.00	219.16	4.6e-05	873	! ecdysone-induced protein E74A +
pir2:S25365	+	188.00	218.23	4.7e-05	966	! CYC8 protein - yeast (Saccharom
pir2:T18235	+	186.50	214.78	6.2e-05	1145	! transcription activator GAL11
pir2:T16870	+	186.00	207.16	7.7e-05	2447	! hypothetical protein T13H2.3 +
pir2:S01992	+	185.50	224.28	5.9e-05	356	! glutenin low molecular weight c
pir2:T18975	-	184.50	230.86	5.9e-05	152	! hypothetical protein C06A1.6 -
pir2:T27431	+	184.00	205.94	0.0001	2129	! hypothetical protein Y79H2A.3
pir2:T05923	+	183.50	221.32	8.2e-05	374	! glutenin low molecular weight c
pir2:T03851	+	183.00	208.39	0.0001	1424	! thyroid hormone receptor activ
pir2:T00661	+	182.50	207.80	0.0001	1417	! hypothetical protein F3I6.24 +
pir2:I46592	+	182.00	220.13	0.0001	347	! involucrin - pig
pir1:JQ1150	+	181.50	213.75	0.0001	648	! protein kinase (EC 2.7.1.37) cA
pir2:T02440	+	181.50	209.84	0.0001	991	! hypothetical protein At2g41450
pir2:T23107	+	181.00	209.62	0.0001	947	! hypothetical protein H20J18.1 +
pir1:S18946	-	180.50	224.87	0.0001	169	! ultra high-sulfur keratin 1 - H
pir2:T06982	+	180.50	217.93	0.0001	359	! glutenin low molecular weight c
pir2:S19033	+	180.50	208.43	0.0002	1008	! transcription activator GAL11
pir1:EDBE23	+	180.00	211.48	0.0002	676	! immediate-early protein IER2
pir2:A53185	+	180.00	211.05	0.0002	708	! G-box-binding factor - slime
pir2:T08588	+	179.00	207.29	0.0002	930	! hypothetical protein L23H3.3
pir2:A56038	+	178.00	205.11	0.0002	1028	! DNA-binding protein ovo - f
pir2:S16356	+	178.00	203.59	0.0002	1213	! ovo protein - fruit fly (Dr

```
seq_name: pir2:T24877

seq_documentation_block:
hypothetical protein T13F2.3a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T24877
R;Swinburne, J.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z19947
A;Accession: T24877
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1074 <WIL>
A;Cross-references: EMBL:Z81122; PIDN:CAB03354.1; GSPDB:GN00022; CESP:T13F2.3a
A;Experimental source: clone T13F2
C;Genetics:
A;Gene: CESP:T13F2.3a
A;Map position: 4
A;Introns: 58/3; 79/1; 112/1; 163/3; 268/1; 388/1; 438/1; 547/3; 628/3; 669/2; 741/1;


alignment_scores:
           Quality:  524.00              Length:    692
             Ratio:  1.444                 Gaps:     24
 Percent Similarity:  52.457   Percent Identity: 25.000

alignment_block:
 US-09-664-641-10 x T24877   ..

 Align seg 1/1  to: T24877  from: 1  to: 1074

    1350 ACAGACTCTTTCAGCACC......TACGAAAAATTTAGAACAGCAGGTGA 1393
          |||   :::|||||||:::      ::::::   ||||||   :::
     347 ThrProThrPheSerSerProAlaHisGlnMetPheArgGlnGlyGlnPr 363

    1394 ATCACAGCCAGCAGGGACATACAAATGCCAATGCAGTGCTGTTTAGCCAA 1443
          :::   |||   |||      :::              :::   |||:
     363 oThrAlaProProGlyMetArgGln.....TyrIleSerProG 376

    1444 GTGAAAGTGACTCCAGAGACACACATGCTACAGCA.....GCAGCA 1484
          ::   ::::::::::::   |||   :::|||          |||
     376 lnAlaGlyGlnGlnGlnProThrProGlyThrProHisArgPheAlaPro 392

    1485 GCAGGCCCAGCAGCAGCAGCAGCACCCGGTTTTACACCTTCAGCCCC 1534
           ||||||:::::::::::::::::::|||   |||:::|||:::||||||
     393 ProGlyProSerGlyThrThrSerSerProHisPheAlaProGlyAlaPr 409

    1535 AGCAGATAAT.................GCAGCTCCAGCAGCAGCAGCAG 1566
          |:::   :::                 :::||| :::|||||||:::
     409 oGlyAlaAspLeuArgSerProSerLeuMetSerPro.AsnGlnGlnSer 425

    1567 CAGCAG..................................... 1572
          ||||||
     426 GlnGlnHisThrProValLeuProArgSerLeuHisAsnGluMetThrSe 442

    1573 .....ATCTCTCAGCAACCTTACCCCCAGCAG. 1599
                :::   :::||||||   ||||||
     442 rHisGlyMetArgProIleAlaLeuProSerGlnProLeuProGlnThrG 459

    1600 .....CCGCCGCAT... 1608
                ||||||:::
     459 lnGlyThrProGlnProIleLeuSerAlaSerSerLeuProProGlnSer 475

    1609 ...........CCATTTTCACAGCAGCAGCAGCAGCAACAGCAGCCACC 1646
                    |||      :::   ||||||   :::|||
     476 AlaGlyAlaThrGlnPheValTyrArgProGlnGlnProAsnGlnMetAs 492
```

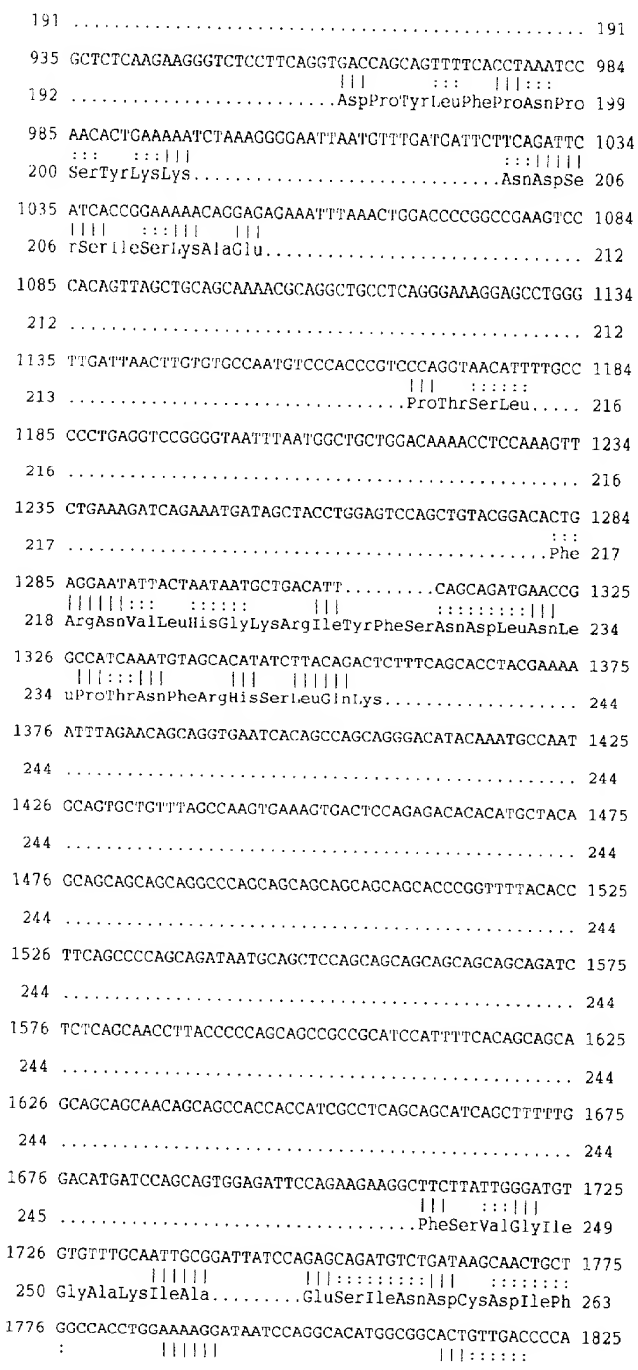




```
seq_name: pir2:T24887

seq_documentation_block:
hypothetical protein T13F2.3b - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T24887
R;Swinburne, J.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z19947
A;Accession: T24887
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
```

```
A;Residues: 1-1076 <WIL>
A;Cross-references: EMBL:Z81122; PIDN:CAB54313.1; GSPDB:GN00022; CESP:T13F2.3b
A;Experimental source: clone T13F2
C;Genetics:
A;Gene: CESP:T13F2.3b
A;Map position: 4
A;Introns: 58/3; 79/1; 114/1; 165/3; 270/1; 390/1; 440/1; 549/3; 630/3; 671/2; 743/1; 90

alignment_scores:
            Quality:  524.00            Length:    692
              Ratio:  1.444               Gaps:     24
 Percent Similarity:  52.457   Percent Identity: 25.000

alignment_block:
 US-09-664-641-10 x T24887   ..

 Align seg 1/1  to: T24887  from: 1  to: 1076

    1350 ACAGACTCTTTCAGCACC......TACGAAAAATTTAGAACAGCAGGTGA 1393
         |||   :::||||||:::      ::::::   ||||||   :::
     349 ThrProThrPheSerSerProAlaHisGlnMetPheArgGlnGlyGlnPr 365

    1394 ATCACAGCCAGCAGGGACATACAAATGCCAATGCAGTGCTGTTTAGCCAA 1443
         :::   |||   |||       :::                :::   |||:
     365 oThrAlaProProGlyMetArgGln.....TyrIleSerProG 378

    1444 GTGAAAGTGACTCCAGAGACACACATGCTACAGCA.........GCAGCA 1484
         ::   ::::::::::::   |||   :::|||             |||
     378 lnAlaGlyGlnGlnGlnProThrProGlyThrProHisArgPheAlaPro 394

    1485 GCAGGCCCAGCAGCAGCAGCAGCAGCACCCGGTTTTACACCTTCAGCCCC 1534
         ||||||:::::::::::::::::::|||   |||:::|||:::|||||
     395 ProGlyProSerGlyThrThrSerSerProHisPheAlaProGlyAlaPr 411

    1535 AGCAGATAAT.................GCAGCTCCAGCAGCAGCAGCAG 1566
         |:::   :::                 :::||| :::||||||:::
     411 oGlyAlaAspLeuArgSerProSerLeuMetSerPro.AsnGlnGlnSer 427

    1567 CAGCAG..................................................... 1572
         ||||||
     428 GlnGlnHisThrProValLeuProArgSerLeuHisAsnGluMetThrSe 444

    1573 .....................ATCTCTCAGCAACCTTACCCCCAGCAG. 1599
                              :::   :::||||||   ||||||
     444 rHisGlyMetArgProIleAlaLeuProSerGlnProLeuProGlnThrG 461

    1600 .....................................CCGCCGCAT... 1608
                                             |||||||:::
     461 lnGlyThrProGlnProIleLeuSerAlaSerSerLeuProProGlnSer 477

    1609 ............CCATTTTCACAGCAGCAGCAGCAGCAACAGCAGCCACC 1646
                     |||      :::   ||||||   :::|||
     478 AlaGlyAlaThrGlnPheValTyrArgProGlnGlnProAsnGlnMetAs 494

    1647 ACCATCGCCTCAGCAGCATCAGCTTTTTGGACATGATCCA.......... 1686
                 |||:::::::|||:::      ||||||
     494 nThrMetAsnGlnMetAsnGlnMetHis...HisAspValAsnProPheP 510

    1687 .....GCAGTGGAGATTCCAGAAGAAGGCTTCTTATTGGGATGTGTGTTT 1731
               :::::::::      |||   :::|||   |||      |||
     510 roIleGlnMetAsnValAspProGluLeuTyrLeuThrGlyIleHisPhe 526

    1732 GCAATTGCGGATTATCCAGAGCAGATGTCTGATAAGCAACTGCTGGCCAC 1781
                |||      :::   ::::::||||||            ::
     527 HisCysPheAspSerGluLysLeuPheGluAspLysLeuAspArgLysAs 543

    1782 CTGGAAAAGGATAATCCAGGCACATGGCGGCACTGTTGAC.........C 1822
         :   :::   :::|||:::   |||||||||   ::::::
     543 nLeuGluPheMetIleLysTyrHisGlyGlyAspIleGluPheGluAlaS 560

    1823 CCACCTTCACGAGTCGA.........TGCACGCACCTTCTCTGTGAGAGT 1863
              |||   :::|||           |||||||:::|||   :::|||

     560 erLysPheHisGluArgIleHisGlnIleThrHisValLeuValAspSer 576

    1864 CAAGTCAGCAGCGCGTATGCACAGGCAATAAGAGAAAGAAAGAGATGTGT 1913
               ::::::            ::::::::::::|||||||||   ||
     577 CysArgAsnThrAsnValMetThrSerLeuGluHisArgLysArgIleVa 593

    1914 TACTGCACACTGGTTAAACACAGTCTTAAAAAAAAAAAAAAATGGTACCGC 1963
         |:::   :::|||:::      |||:::::::::::::::::::      |
     593 lSerIleGlnTrpIleIleAspValMetGluArgGluArgValGluLeuP 610

    1964 CGCACCGAGCCCTTCACTTCCCAGTGGCCTTCCCACCAGGAGGAAAGCCA 2013
         ||   |||       |||:::|||       |||       :::   :::|||
     610 roTrpArgIleAlaHisLeuProProProPheAsnGluAsnPheArgPro 626

    2014 TGTTCACAGCATATTATTTCTGTGACTGGATTTGTTGATAGTGACAGAGA 2063
                    ::::::|||::::::::|||||||   :::|||:::|||
     627 TyrLeuGlyLysLeuPheSerLeuSerGlyPheAspSerSerGluArgAl 643

    2064 TGACCTAAAATTAATGGCTTATTTGGCAGGTGCCAAATATACGGGTTATC 2113
         ::::::::::||||||   :::   ||||||||||   |||   :::|
     643 aAlaIleSerPheMetAlaGluMetMetGlyAlaLysIleThrProPheL 660

    2114 TATGCCGCAGCAACACAGTCCTCATCTGTAAAGAACCAACTGGTTTAAAG 2163
         ||:::|||:::|||   :::|||||||:::|||:::|||:::      |||
     660 euAlaArgGlnAsnAspLeuLeuIleAlaLysGlnProAlaAsp...Lys 675

    2164 TATGAAAAAGCCAAAGAGTGGAGGATACCCTGTGTCAACGCCCAGTGGCT 2213
               :::|||:::||||||::::::|||   ||||||   ||||||::
     676 ValThrArgAlaGlnGluTrpLysValProValValAsnPheGlnTrpIl 692

    2214 TGGCGACATTCTTCTGGGAAACTTTGAGGCACTGAGGCAG.......... 2253
         ::::|||       :::   :::   ||||||       |||
     692 eAlaAspAlaTyrValCysAsp...GluAlaSerProGlnLysArgProA 708

    2254 ..ATTCAGTATAGTCGCTACACGGCATTCAGTCTGCAGGATCCATTTGCC 2301
           ::::::       ||||||          :::|||:::::|||
     708 snValGluAsnProArgTyr.........GlnLeuGluAsnProValMet 721

    2302 CCTACCCAGCATTTAGTTTTAAAT.......................... 2325
            :::   :::          |||
     722 ThrAsnValAsnThrSerProAsnHisIleGluGlyAlaSerGlnGluPh 738

    2326 .......CTTTTAGATGCTTGGAGAGTTCCCTTAAAAGTGTCTGCAGAGT 2368
                :::::::::::::|||:::                ::::::   :::
     738 eGlyAlaMetMetSerCysTrpLysSerGlnThrIleIleAlaAspGlnA 755

    2369 TGTTGATGAGTATAAGACTACCTCCCAAACTGAAACAGAATGAAGTA... 2415
                  :::   |||       |||       ||||||::::::
     755 laTyrGluLysAlaArgGlnAsnArgLysThrLeuGlnAsnAsnThrPhe 771

    2416 ............................GCTAATGTCCAGCCTTCT..... 2433
                                       |||   |||:::
     772 PhePheProSerLysArgLeuSerAspGlnAlaValAlaProThrGluAs 788

    2434 ..........TCCCAAAGAGCCAGAATTGAAGACGTA.............. 2460
                   |||:::|||||||||::::::::::::
     788 pGluIleLysSerSerArgAlaArgLeuAspGluIleLysThrHisTrpG 805

    2461 ........................CCACCTCCCACTAAAAAGCTA... 2481
                                    |||   :::|||::::::
     805 luGluHisGlnGluLysMetLysAspThrProLeuSerLysGluPheTyr 821

    2482 .........ACTCCAGAATTGACCCCTTTTGTGCTTTTC...ACTGGATT 2519
                   |||:::            :::   |||      |||::
     822 ArgValArgGluProGlnAlaArgValSerAlaTrpPheGlyGluGlyIl 838

    2520 CGAGCCTGTCCAGGTTCAACAGTATATTAAGAAGCTCTACATTCTTGGTG 2569
         ::::         :::                |||||||||   :::|||||||
     838 eAspAspGluAlaLeuThrIleLeuLysLysLysLeuGluPheLeuGlyG 855

    2570 GAGAGGTTGCGGAGTCTGCACAGAAGTGCACACACCTCATTGCCAGCAAA 2619
         |||||   :::|||:::   :::   :::||||||::|||      :::
     855 lyGluCysValGluLysIleArgAspAlaThrHisValIleLeuIleSer 871
```

```
 2620 GTGACTCGCACCGTGAAGTTCCTGACGGCGATTTCTGTCGTGAAGCACAT 2669
          |||::::::   :::|||   :::|||          |||:::||
  872 GlyArgArgSerLeuValLeuLeuGluSerIleIleArgGlyLysAsnIl 888

 2670 AGTGACGCCCAGAGTGGCTGGAAGAATGCTTCAGGTGTCAGAAGTTCATTG 2719
      |:::   ||||||||||:::   :::   ::::::   :::   :::::::|
  888 eMetAspProGluTrpIleValAspSerTyrLysGlnLysMetTrpLeuA 905

 2720 ATGAGCAGAACTACATTCTCCGAGATGCTGAGGCAGAAGTACTTTTCTCT 2769
      ||      :::|||:::|||:::|||   |||   |||      |||:::
  905 spThrLeuAspTyrPheLeuHisAspLysGluLeuGluLysGluPheAla 921

 2770 TTCAGCTTGGAAGAATCCTTAAAACGGGCACACGTTTCTCCACTCTTTAA 2819
      ::::::   ::::::|||:::   ||||||:::   :::|||:::|||::
  922 TyrAsnCysLysArgSerValLeuArgAlaArgAsnLysProValPheGl 938

 2820 GGCAAAATATTTTTACATCACACCTGGAATCTGCCCAAGTCTTTCCACTA 2869
      :         |||::::::|||      :::   |||:::   :::   :
  938 uAspIleGluPheHisValThrArgPheValGluProAsnGlnAsnAspL 955

 2870 TGAAGGCAATCGTAGAGTGTGCAGGAGGAAAGGTGTTATCCAAGCAGCCA 2919
      ::      :::::::|||   :::||||||:::|||   |||:::::::|||
  955 euValArgLeuIleGluLeuGlyGlyGlyAsnValHisSerGluLysPro 971

 2920 TCT.........TTCCGGAAGCTCATGGAGCACAAGCAGAACTCGAGTTT 2960
      :::         :::   |||   :::|||   :::|||
  972 AspProLysTyrLeuAlaLysCysValGluThrGluGln.......... 984

 2961 GTCGGAAATAATTTTAATATCCTGTGAAAATGACCTTCATTTATGCCGAG 3010
            :::|||:::|||||||||||||||||||   ::::::
  985 ....ProPheIleIleIleSerCysGluAsnAspAlaArgPheLeuSerT 1000

 3011 AATATTTTGCCAGAGGCATAGATGTTCACAATGCAGAGTTCGTTCTGACT 3060
            :::      :::   :::::::|||:::::::::|||||||
 1000 yrLeuAlaGlySerLysLeuProIleTyrAsnValAspLeuValLeuPhe 1016

 3061 GGAGTGCTCACTCAAACGCTGGAC 3084
      ::::::|||   |||   ::::::
 1017 AlaMetLeuArgGlnGlnIleGlu 1024

seq_name: pir2:T37978

seq_documentation_block:
multiple BRCT domain containing protein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C;Accession: T37978; T40564
R;Murphy, L.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, March 1996
A;Reference number: Z21759
A;Accession: T37978
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-878 <MUR>
A;Cross-references: EMBL:Z69909; PIDN:CAA93789.1; GSPDB:GN00066; SPDB:SPAC19G10.07
A;Experimental source: strain 972h-; cosmid c19G10
R;Wood, V.; Rajandream, M.A.; Barrell, B.G.; Davis, P.; Churcher, C.M.
submitted to the EMBL Data Library, July 1999
A;Reference number: Z21937
A;Accession: T40564
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-878 <WOO>
A;Cross-references: EMBL:AL096788; PIDN:CAB46668.1; GSPDB:GN00067; SPDB:SPBC582.05c
A;Experimental source: strain 972h-; cosmid c582
C;Genetics: <GEN1>
A;Gene: SPAC19G10.07
A;Map position: 1
A;Introns: 37/3; 83/3
C;Genetics: <GEN2>
A;Gene: SPBC582.05c
A;Map position: 2

alignment_scores:
            Quality:  314.50              Length:   1194
              Ratio:  0.758                 Gaps:     35
Percent Similarity:  34.757   Percent Identity: 17.169

alignment_block:
US-09-664-641-10 x T37978   ..

Align seq 1/1  to: T37978  from: 1  to: 878

  144 TACGGATCTCTCGGCCCCTGCCCTTCAAATCACCTCTAGGCACAGGTTAT 193
      |||:::::::|||  |||      :::|||:::|||      :::  ::
   25 TyrSerAlaLeuGlnPro...AsnGlnAsnGluLeuArgLysLysGluLe 40

  194 TCAGCTTCTCAAGGCTGGAAAAGCGAAGGAAGTTTCCTACAAT...GCAC 240
      :  :::  :::  |||||||||  :::      |||:::      |
   40 uPheIleLysAsnAspGlyLysAlaLeuSerPheProTyrAspTrpLysL 57

  241 TAGCCTCACACATAATCTCAGAGGATGGGGACAATCCAGAGGTGGGAGAA 290
      |||||:::|||:::|||  :::|||  ::::::|||:::|||  |||
   57 euAlaThrHisValIleCysAspAspPheSerSerProAsnValGlnGlu 73

  291 GCTCGGGAAGTCTTTGACTTACCTGTTGTAAAGCCTTCTTGGGTGATTCT 340
      :::  :::      :::|||  :::::::|||  :::|||:::
   74 GlySerLys...ArgSerLeuArgLeuAlaLysThrAsnTrpIleArgAs 89

  341 GTCCGTTCAGTGTGGAACTCTTCTGCCAGTAAATGGTTTTTCTCCAGAAT 390
          |||:::  :::|||||||||||      :::  :::|||  :::
   89 pCysValAspLysAsnThrLeuLeuAsnTyrSerPheTyrSerCysAsnP 106

  391 CATGTCAGATTTTTTTTGGAATCACTGCCTGCCTTTCTCAGGGTGTTGAT 440
            ::|||  ||||||  |||          |||  :::|||
  106 roTyrLeuLeuPheLysGlyIleCysAlaSerSerCysGln...IleAsp 121

  441 ACAAGCTGGAGCTCTTTGTTGGAGTCTTCCAGAGCTCTCCCAAGGAGAGG 490
      :::      ||||||||||::::::
  122 SerTyrGlnSerSerLeuIleAsp..........................  129

  491 TAGGGAAGGGAGCTTGTCCAGCAGAAGTTGGGAAGCACAGAGATCATCTG 540

  129 ..........................................................  129

  541 CCTTCTTCTGACCCGGTATTGATGCAGGCTGAGGCCTCTGTTGTAATGTG 590
                        ||||||    |||:::
  130 ...................AspAla.LeuGluThr.............  134

  591 CTGGGTGTCATCTGAAGACAGAAGTGCCCTGTGGGCTTTGGTTACGTTCT 640

  135 ..........................................................P  135

  641 ATGGGGGAGATTGCCAGCTAACCCTCAATAAGAAATGCACGCATTTGATT 690
      ::||||||      :::      |||  |||:::  ||||||||||:::
  135 heGlyGlyArgPheSerLysGlyLeuMetLysSerMetThrHisLeuPhe 151

  691 GTTCCAGAGCCAAAGGGGGGGAAATACGAATGTGCTTTAAAGCGA..... 735
      :::   :::      |||::|||  :::   :::|||  :::
  152 ThrTyrSerGlyMetGlyAlaLysCysLysLysValLeuAspLysProSe 168

  736 .GCAAGTATTAAAATTGTGACTCCTGACTGGGTTCTGGATTGCGTATCAG 784
       ||||||||||::::::  |||:::|||:::||||||||||::::::
  168 rLeuSerIleLysLeuIleHisProGlnTrpLeuLeuAspCysLeuGlnP 185

  785 AGAAAACCAAAAAGGACGAAGCATTTTATCATCCTCGTCTGATTATTTAT 834
                |||:::
  185 heGlyGlnLeuIleAspGln.............................  191

  835 GAAGAGGAAGAAGAGGAAGAGGAAGAGGAGGAGGAAGTAGAAAATGAGGA 884

  191 ..........................................................  191

  885 ACAAGATTCTCAGAATGAGGGTAGTACAGATGAGAAGTCAAGCCCTGCCA 934
```



```
2386 .........................CTACCTCCCAAACTGAAACAGAAT 2409
                              |||      ||||||   :::|||
 545 laSerSerAspAspGluLeuProValLeuAlaThrLysLeuValAspAsn 561

2410 GAAGTAGCTAATGTCCAGCCTTCTTCCCAAAGAGCCAGAATT........ 2451
       :::   :::   :::|||   |||          ::::::
 562 ValIleLysGluLysSerProLeuSerLeuThrProLysValValValPr 578

2452 .......GAAGACGTACCACCTCCCACTAAAAAGCTAACTCCAGAATTG. 2493
             :::::::::           |||||||||      ||||||
 578 oSerHisLysGluThrTyrThrAspGluLysLysLeuIleAspGluLeuA 595

2493 ............................................... 2493

 595 spArgValAsnProLeuAsnSerSerGlnLeuLeuArgSerLysArgLys 611

2493 ............................................... 2493

 612 SerAlaAlaThrAlaLeuSerMetLeuGlnAsnValIleMetProAspVa 628

2493 ............................................... 2493

 628 lLeuAlaPheGluArgGluLysLysArgArgGlnThrHisArgSerValS 645

2493 ............................................... 2493

 645 erSerGlyGluValSerArgGluSerSerGluSerArgAsnThrAsnAla 661

2494 ............ACCCCTTTTGTGCTTTTCACTGGATTCGAGCCTGTCCA 2531
                   ::::::   |||||||||:::::
 662 LysAlaSerLysArgValTyrIleThrPheThrGlyTyrAspLysLysPr 678

2532 CGTTCAACAGTATATTAAGAAGCTCTACATTCTTGGTGGAGAGGTTCCGG 2581
       :::   :::|||||||||   :::          ::::::::::
 678 oSerIleAspAsnLeuLysLysLeuAspMet.....SerIleThrS 692

2582 AGTCTGCACAGAAGTGCACACACCTCATTGCCAGCAAAGTGACTCGCACC 2631
     :::::   :::|||||||||||||||||||||   ::::::   ||||||
 692 erAsnProSerLysCysThrHisLeuIleAlaProArgIleLeuArgThr 708

2632 GTGAAGTTCCTGACGGCGATTTCTGTCGTGAAGCACATAGTGACGCCAGA 2681
       |||||||||   :::|||              :::|||||||   ::
 709 SerLysPheLeuCysSerIleProTyrGlyProCysValValThrMetAs 725

2682 GTGGCTGGAAGAATGCTTCAGCTGTCAGAAGTTCATTGATGAGCAGAACT 2731
     :|||::::::::::|||:::::   :::::::::::::||||||:::   |
 725 pTrpIleAsnSerCysLeuLysThrHisGluIleValAspGluGluProT 742

2732 ACATTCTCCGAGATGCTGAGGCAGAAGTACTTTTCTCTTTCAGCTTGGAA 2781
     ||:::|||:::|||   |||   |||:::   ::::::   :::||||||
 742 yrLeuLeuAsnAspProGluLysGluLeuGluLeuGlyCysThrLeuGlu 758

2782 GAATCCTTAAAACGGGCACACGTTTCT......CCACTCTTTAAGGCAAA 2825
     ::::::|||||||||||||:::::::::        |||::::::
 759 SerAlaLeuLysArgAlaArgAlaGlnGlyProSerLeuLeuGluAspTy 775

2826 ATATTTTTACATCACA...CCTGGAATCTGCCCA...AGTCTTTCCACTA 2869
            |||:::|||         :::::::|||   ::::::   ::::
 775 rValValTyrLeuThrSerLysThrValAlaProGluAsnValProAlaV 792

2870 TGAAGGCAATCGTAGAGTGTGCAGGAGGAAAGGTGTTATCCAAGCAGCCA 2919
     ::   :::||||||::      ||||||         :::   :::
 792 alIleSerIleValLysSerAsnGlyGlyValCysSerThrLeuAsnVal 808

2920 TCTTTCCGGAAGCTCATGGAGCACAAGCAGAACTCGAGTTTGTCGGAAAT 2969
           :::::::|||   :::|||   ::::::::::::          ::
 809 TyrAsnLysArgLeuAlaArgHisLeuGluAspGlyAsn.........Va 822

2970 AATTTTAATATCCTGTGAAAATGACCTTCATTTATGCCGAGAATATTTT. 3018
     ::::|||||||:::|||::::::|||   |||:::      :::::::::
 822 lValLeuLeIleThrCysAsnGluAspSerHisIleTrpThrAsnPheLeuA 839

3019 .............GCCAGAGGCATAGATGTTCACAATGCAGAGTTCGTT 3054
                    :::   |||   :::::::|||   :::::::::
 839 spAsnAlaSerGlnAsnLysThrIlePheLeuGlnAsnTyrAspTrpLeu 855

3055 CTGACTGGAGTGCTCACTCAAACGCTGGAC 3084
       :::      ||||||   |||   :::|||
 856 IleLysThrValLeuArgGlnGluIleAsp 865
```

seq_name: pir2:G85039

seq_documentation_block:
hypothetical protein AT4g03130 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
C;Accession: G85039
R;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sp
Nature 402, 769-777, 1999
A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A;Reference number: A85001; MUID:20083488
A;Accession: G85039
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-765 <STO>
A;Cross-references: GB:NC_001268; NID:g7270183; PIDN:CAB77798.1; GSPDB:GN00140
C;Genetics:
A;Gene: AT4g03130
A;Map position: 4

alignment_scores:

Quality:	265.50	Length:	910
Ratio:	0.676	Gaps:	34
Percent Similarity:	43.187	Percent Identity:	19.890

alignment_block:
US-09-664-641-10 x G85039 ..

Align seg 1/1 to: G85039 from: 1 to: 765

```
 766 GTTCTGGATTGC......GTATCAGAGAAAACCAAAAAGGACGAAGCATT 809
     :::|||   |||      :::::::::           ||||||||||::
  26 IleLeuGlyCysGlyGluIleGluAspThrGluAlaIleAspGluAlaLe 42

 810 TTAT.....................CATCCTCGTCTGATTATTTATGAAG 838
     :                        |||      ::::::|||::::
  42 uValLeuAsnSerSerLeuLeuGluThrProPheGlnValLeuTyrAspA 59

 839 AGGAAGAA.....................GAGGAAGAGGAAGAGGAG 864
     ::   |||                   :::|||||||:::   :::
  59 spThrGluLeuValAspAsnGlyAsnCysGlyAspGluGluSerLeuAsn 75

 865 GAGGAAGTAGAAAATGAGGAACAA.....GATTCTCAGAATGA 902
     :::|||      :::   :::|||          ||||||::::::||
  76 GlnGluCysTyrGlyLeuAspGlnValValValAspAspSerAspAspGl 92

 903 GGGTAGTACAGATGAGAAGTCAAGCCCTGCCAGCTCTCAAGAAGGGTCTC 952
     |   ::::::         ::::::|||          |||   ||||
  92 uAspGlySerValAlaAlaGlyGluProMetCysLeuProGluLeuSerP 109

 953 CT...TCAGGTGAC..................................... 963
     ||   :::|||||||
 109 roLysAlaGlyAspIleLeuLeuGluSerAspGlySerAsnAspHisGlu 125

 964 ...CAGCAGTTTTCACCTAAATCCAAC..................... 987
        |||      :::   :::|||||||
 126 CysGlnThrGlyLysGlnGluSerAsnCysAspThrValThrGlyPheGl 142

 988 ...............ACTGAAAAATCTAAAGGGGAA...TTAATGTTTG 1018
                     |||   :::|||   |||      |||   |||
 142 nGlySerLeuArgIleThrAlaSerSerHisGlyProGlyLeuAspPheL 159

1019 ATGATTCTTCAGATTCATCACCGGAAAAACAG.................. 1050
       ||||||::::::   :::   :::   |||
```

```
 159 euAspSerGlnGluProGlyGluSerThrGlnAlaAspAlaLeuGlyPhe 175

1051 ..................GAGAGAAATTTAAACTGGACCCCGGCCGAAGT 1082
                       :::::::::||||||   :::|||:::::::
 176 ValAspHisPheLeuMetAspLysAspLeuAsnLeuSerProValAspLe 192

1083 CCCACAGTTAGCTGCAGCAAAACGCAGGCTGCCTCAGGGAAAGGAGCCTG 1132
     :|||      ::::::   ::::::      |||   :::         |
 192 uProValAsnSerCysArgArgLysSerSerProValSerGlyAlaLysG 209

1133 GGTTGATTAACTTGTGTGCCAATGTCCCACCCGTCCCAGGTAACATTTTG 1182
     ||      :::|||:::   ::::::
 209 lyCysGlnSerLeuAlaLysArgIleMetSer..................  219

1183 CCCCCTGAGGTCCCGGGTAATTTAATGGCTGCTGGACAAAACCTCCAAAG 1232
                                                  :::||
 220 ..............................................ArgSe 221

1233 TTCTGAAAGATCAGAAATGATAGCTACCTGG..................  1263
     |      |||:::      :::      |||
 221 rProThrArgLysMetSerValPheAspTrpSerSerAspGlnCysAspV 238

1264 .................AGTCCAGCTGTACGG.........ACACTGAGG 1287
                      |||||||:::::|||         |||
 238 alSerAspProLysThrSerProValThrArgAlaSerIleThrCysSer 254

1288 AATATTACTAATAATGCT............GACATTCAGCAGATGAACCG 1325
     :::|||   ::::::                |||:::         :::||
 255 LysIleArgGluAspGlnValAlaGlyAspAspPheGlyValLysLysAr 271

1326 GCCATCAAATGTAGCACATATCTTACAGACTCTTTCAGCACCTACGAAAA 1375
     |      :::   :::   |||   ::::::   |||   ||||||:::::
 271 gSerArgAspProCysAspIleArgLysValSerSerHisProThrGlnA 288

1376 ATTTAGAACAGCAGGTGAATCACAGCCAGCAGGGACATACAAATGCCAAT 1425
     :::::   :::                          :::::::::|||:::
 288 rgValMetLys......................AsnSerSerAlaLys 296

1426 GCAGTGCTGTTTAGCCAAGTGAAAGTGACTCCAGAGACACACATGCTACA 1475
        :::   ::::::|||:::::         |||      |||:::
 297 HisLeuLysMetGluGlnAlaSerGlyLeuSerGluGlyIleMetPheIl 313

1476 GCAGCAGCAGCAGGCCCAGCAGCAGCAGCAGCACCCGGTTTTACACC 1525
      :::|||:::::|||||   ::::::::      |||         |
 313 eSerGlnLysAspAlaGlnLeuGluAspLysAla...ProLysGluProL 329

1526 TTCAGCCCCAGCAGATAATG.....................CAGCTCCAG 1554
     ||:::|||::::::   :::                    ||':::
 329 euGluProGluGluAspPheIleAspIleGlyIleAsnThrGlnIleAla 345

1555 CAGCAGCAGCAGCAGATCTCTCAGCAACCTTACCCC.............  1593
        :::   ::::::   :::      |||
 346 AlaGluAlaMetSerAlaLeuValTyrAlaProCysThrThrGluGluAl 362

1594 ....CAGCAGCCGCCGCATCCATTTTCACAGCAGCAGCAGCAGCAACAGC 1639
         ::::::   |||   |||   |||   :::::::::::::::|||
 362 aCysGluSerAspProIleProGlySerValSerGluMetArgAspGlnV 379

1640 AGCCACCACCATCGCCTCAGCAGCATCAGCTTTTTGGACATGATCCAGCA 1689
                |||   ::::::::::::   :::      :::|||
 379 alSerAsnLeuSerArgArgAsnAsnAspThrIleGluGlyGluPro... 394

1690 GTGGAGATTCCAGAAGAAGGCTTCTTATTGGGATGTGTGTTTGCAATTGC 1739
                     |||         |||:::   :::|||:::   ||
 395 ..............GluArgAspLysLeuAsnGlyLeuPheSerAlaAl 406

1740 GGATTATCCAGAGCAGATGTCTGATAAGCAACTGCTGGCCACCTGGAAAA 1789
     |   :::   |||:::   |||   |||:::         |||   ::::
 406 aProHisArgGluArgAsnSerLysLysLysArgLysPheThrAsnGluG 423

1790 GGATAATCCAGGCACATGGCGGCACTGTTGACCCCACCTTCACGAGTCGA 1839
     ::      :::|||:::                    :::   ::::::
 423 luArgThrArgAlaAsn.................ValSerValLysThr 433

1840 TGCACGCACCTTCTCTGTGAGAGTCAAGTCAGCAGCGCGTATGCACAGGC 1889
     |||         |||||||||   :::            :::   :::||
 434 CysLeuLeuAsnLeuCysGluTrpArg............HisProArgAl 446

1890 AATAAGAGAAAGAAAGAGATGTGTTACTGCACACTGGTTAAACACAGTCT 1939
     |   |||:::|||                                    |
 446 aLysArgSerArg.....................................L 451

1940 TAAAAAAAAAAAAAATGGTACCGCCGCACCGAGCCCTTCACTTCCCAGTG 1989
     ||   ::::::      ||||||||||:::|||:::
 451 euMetGlnArgHisHisValProProArgArgSerTrp............ 463

1990 GCCTTCCCACCAGGAGGAAAGCCATGTTCACAGCATATTATTTCTGTGAC 2039

 463 ................................................. 463

2040 TGGATTTGTTGATAGTGACAGAGATGACCTAAAATTAATGGCTTATTTGG 2089

 463 ................................................. 463

2090 CAGGTGCCAAATATACGGGTTATCTATGCCGCAGCAACACAGTCCTCATC 2139
       |||||:::   :::            :::::|||||
 464 ..GlyAlaSerLeuValLysAspArgSerGluThrAsnThr..... 476

2140 TGTAAAGAACCAACTGGTTTAAAGTATGAAAAAGCCAAAGAGTGGAGGAT 2189

 476 ................................................. 476

2190 ACCCTGTGTCAACGCCCAGTGGCTTGGCGACATTCTTCTGGGAAACTTTG 2239
                          |||:::::::   |||   |||:::::::
 477 ....................LeuSerSerSerLeuArgGlySerLeuS 486

2240 AGGCACTGAGGCAGATTCAGTATAGTCGCTACACGGCATTCAGTCTGCAG 2289
     :::::   ||||||
 486 erGlyThrAr GlnAla.................................. 491

2290 GATCCATTTGCCCCTACCCAGCATTTAGTTTTAAATCTTTTAGATGCTTG 2339
              :::      |||      |||:::::::|||      |||::
 492 .........SerSerCysGlnSerGlyValIleAspLeuAsnValAlaTy 505

2340 GAGAGTT...CCCTTAAAAGTG.............................. 2358
     :::::::   |||   |||:::
 505 rHisAlaSerProArgLysIleTyrAspGlySerHisGluSerProCysA 522

2359 ..........................TCTGCAGAGTTGTTGATGAGTATA 2382
                               :::   ||||||     |||:::
 522 snLysAspPheProArgLeuPheLeuGlnLysGluLeuThrThrSerLeu 538

2383 AGACTACCTCCCAAACTG.........................AAACA 2405
           |||   |||:::                         |||::
 539 GlyGlyProGlyLysIleGlyAspPheValTrpLysAspLeuArgLysAr 555

2406 GAATGAAGTAGCTAATGTCCAGCCTTCTTCCCAAAGAGCCAGAATTGAAG 2455
     ::::::::::|||:::|||:::
 555 gArgAsnLeuAlaHisValArg............................ 562

2456 ACGTACCACCTCCCACTAAAAAGCTAACTCCAGAATTGACCCCTTTTGTG 2505
                                                     |||
 563 ..............................................Val 563

2506 CTTTTCACT...GGATTCGAGCCTGTCCAGGTTCAACAGTATATTAAGAA 2552
     ||||||:::   :::::::::         ||:::|||      |||
 564 LeuPheSerGlnAsnLeuAspAspGluThrValLysGlnGlnLysLysIl 580

2553 GCTCTACATTCTTGGTGGAGAGGTTGCGGAGTCTGCACAGAAGTGCACAC 2602
      :::      ||||||   :::   |||:::|||:::          ||||
 580 eMetValArgLeuGlyIleSerProAlaSerSerSerAlaAspSerThrH 597

2603 ACCTCATTGCCAGCAAAGTGACTCGCACCGTGAAGTTCCTGACGGCGATT 2652
     ||:::||||||::::::   :::||||||   :::::::|||   ||||||
 597 isPheIleAlaAspArgPheAlaArgThrArgAsnMetLeuGluAlaIle 613
```

```
 2653 TCTGTCCTGAAGCACATAGTGACGCCAGAGTGGCTGGAAGAATGCTTCAG 2702
 614 AlaLeuGlyLysPheValValThrProIleTrpLeuGluSerCysAlaGl 630

 2703 GTGTCAGAAGTTCATTGATGAGCAGAACTACATTCTCCGAGATGCTGAGG 2752
 630 nThrArgCysLeuIleAspGluLysSerTyrIleLeuArgAspIleLysL 647

 2753 CAGAAGTACTTTTCTCTTTCAGCTTGGAAGAATCCTTAAAACGGGCACAC 2802
 647 ysGlu...LysAspGlyPheCysLeuLeuThrSerLeuAlaArgAlaLys 662

 2803 GTTTCTCCACTCTTTAAGGCAAAATATTTTTACATCACACCTGGAATCTG 2852
 663 GlnHisProLeuLeuLysGlyPheLysValCysIleThrProSerIleLy 679

 2853 CCCAAGTCTTTCCACTATGAAGGCAATCGTAGAGTGTGCAGGAGGAAAGG 2902
 679 sProSerArgGlyMetIleThrAspLeuValLysMetThrGlnGlyGlnV 696

 2903 TGTTATCCAAGCAGCCATCTTTCCGGAAGCTCATGGAGCACAAGCAGAAC 2952
 696 alValGluAlaSer.....GluIleIleAlaAlaGluAspArg 708

 2953 TCGAGTTTGTCGGAAATAATTTTAATATCCTGTGAAAATGACCTTCATTT 3002
 709 AsnPheProGluAspValLeuIleLeuSerCysLysGluAspArgAspPh 725

 3003 ATGCCGAGAATATTTTGCCAGAGGCATA..... 3030
 725 eCysLeuProPheValAsnGlnGlyThrValSerSerGlnPheGlnArgL 742

 3031 .....GATGTTCACAATGCAGAGTACCTTCTGACTGGAGTGCTC 3069
 742 ysGlyThrLysGluIleHis.....LysPheThrAlaAlaCysLeuLeu 756

 3070 ACTCAAACGCCTGGACTATGAATCATATAAG 3099
 757 LeuSerThrThrAspTrp.....TyrLys 764

seq_name: pir2:T49721

seq_documentation_block:
hypothetical protein B23L21.340 [imported] - Neurospora crassa
C;Species: Neurospora crassa
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
C;Accession: T49721
R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Partmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, May 2000
A;Reference number: Z25022
A;Accession: T49721
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-831 <SCH>
A;Cross-references: EMBL:AL356172; GSPDB:GN00116; NCSP:B23L21.340
A;Experimental source: BAC clone B23L21; strain OR74A
C;Genetics:
A;Gene: NCSP:B23L21.340
A;Map position: 6
A;Introns: 31/3; 629/3


alignment_scores:
            Quality:  229.00              Length:    725
              Ratio:  0.727                 Gaps:     21
 Percent Similarity:  43.448   Percent Identity: 18.207

alignment_block:
 US-09-664-641-10 x T49721   ..

 Align seg 1/1  to: T49721  from: 1  to: 831

    1156 GTCCCACCCGTCCCAGGTAACATTTTGCCCCCTGAGGTCCGGGGTAATTT 1205

      71 IleProValValAsnSerAsnTrpIleLysIleThrLeuAlaArgAsnLy 87

    1206 AATGGCTGCTGGACAAAACCTCCAAAGTTCTGAAAGATCAGAAATGATAG 1255
      87 sValAla.....GlnValA

    1256 CTACCTGGAGTCCAGCTGTACGGACACTGAGGAAT.....ATTACTAAT 1299
      92 rgProTyrSerProAspProArgMetIlePheSerAsnValIleLeuThr 108

    1300 AATGCTGACATTCAGCAGATGAACCGGCCATCAAATGTAGCACATATCTT 1349
     109 CysAlaAspIleProGluSerAspLysGluThrIleIleGlyAlaThrMe 125

    1350 ACAGACTCTTTCAGCACCTACGAAAAATTTAGAACAGCAGGTGAATCACA 1399
     125 tAlaLeuGlyGlyMetGluSerLysAspLeuThrArgGlnThrThrHisI 142

    1400 GCCAGCAGGGACATACAAATGCCAATGCAGTGCTGTTTAGCCAAGTGAAA 1449
     142 le..... 142

    1450 GTGACTCCAGAGACACACATGCTACAGCAGCAGCAGCAGGCCCAGCAGCA 1499
     143 .....CysAlaLeuSerMetAspHisGluLysCysGlnGluAl 155

    1500 GCAGCAGCAGCACCCGGTTTTACACCTTCAGCCCCAGCAGATAATGCAGC 1549
     155 aGlnLysLysAsnProLysCysLysIleValLeuProHisTrpPheAspA 172

    1550 TCCAGCAGCAGCAGCAGCAGCAGATCTCTCAGCAACCTTACCCCCAGCAG 1599
     172 spCysPheArgLeuGlyArgArgIleSerGluAlaProTyrMetLeu... 187

    1600 CCGCCGCATCCATTTTCACAGCAGCAGCAGCAGCAACAGCAGCCACCACC 1649
     188 .....Pr 188

    1650 ATCGCCTCAGCAGCATCAGCTTTTTGGACATGATCCAGCAGTGGAGATTC 1699
     188 oAsnProGluIleLeuArgThr...GlyProAspGluAspIleThrIleP 204

    1700 CAGAAGAAGGCTTCTTATTGGGATGTGTGTTTGCAATT.....GCGGAT 1743
     204 roProSerGluAlaValGluGlyAlaThrSerValIleProSerAlaAla 220

    1744 TATCCAGAGCAGATGTCTGATAAGCAACTGCTGGCCACCTGGAAAAGGAT 1793
     221 MetProGlyArgGluArgGluLysIleValValPheAlaGlnLysLysVa 237

    1794 AATC..... 1797
     237 lMetIleSerGlnAspLeuProIleAsnAlaArgLeuArgAsnIleLeuA 254

    1798 .....CAGGCACATGGCGGCACTGTTGACCCCACCTTCACG 1833
     254 snAsnLeuIleGluGlnSerAspGlyGluValValAsnAsp.....Val 268

    1834 AGTCGATGCACGCACCTTCTCTGTGAGAGTCAAGTCAGCAGCGCGTATGC 1883
     269 AspAlaCysAspMetTyrValCysGlnTyrArgAspGlyAspGluTyrIl 285

    1884 ACAGGCAATAAGAGAAAGAAAGAGATGTGTTACTGCACACTGGTTAAACA 1933
     285 eArgAlaAlaGlnGlnGlyLysAspValGlyAsnLeuAlaTrpLeuTyrT 302

    1934 CAGTCTTAAAAAAAAAAAAAATGGTA...CCGCCGCACCGAGCCCTTCAC 1980
     302 yrLeuIleValHisAsnGluTrpThrArgProThrArgArgLeuLeuHis 318

    1981 TTCCCAGTGGCCTTCCCACCAGGAGGAAAGCCATGTTCACAGCATATT.. 2028
```

```
 319 TyrProIle......ProArgAspGlyIleProGlyPheThrGlyMetLy 333

2029 .ATTTCTGTGACTGGATTTGTTGATAGTGACAGAGATGACCTAAAATTAA 2077
      |||:::::::::::::::::      :::   |||      |||:::   :
 333 sIleThrLeuSerAsnTyrGlyGlyAspAlaArgValTyrLeuGluAsnL 350

2078 TGCCTTATTTGGCAGGTGCCAAATATACGGGTTATCTATGCCGCAGCAAC 2127
      ::         |||||||||   ||||||      :::      :::|||
 350 euIleThrAlaAlaGlyAlaThrTyrThrLysThrMetLysAlaAspAsn 366

2128 ACAGTCCTCATCTGTAAAGAACCAACTGGTTTAAAGTATGAAAAAGCCAA 2177
      |||   ||||||      :::   :::|||   |||   |||   |||||
 367 ThrHisLeuIleThrAlaArgMetAsnGlyGluLysCysGluAlaAlaLy 383

2178 AGAGTGGAGGATACCCTGTGTCAACGCCCAGTGG..............  2211
      |||||||:::|||      :::|||     |||
 383 sGluTrpAsnIleGluIleIleAsnHisLeuTrpIleGluAspSerTyrA 400

2211 ..................................................  2211

 400 laLysCysGluValGlnThrPheSerAsnProArgTyrGlnHisPhePro 416

2212 ...........CTTGGCGACATTCTTCTGGGAAACTTT.........GA 2240
                |||||||:::|||:::   :::::|||          ::
 417 ProArgThrAsnLeuGlyGluIleIleGlySerThrPhePheAspGluGl 433

2241 GGCACTGAGGCAGATTCAGTAT.......................... 2262
      :   |||||:::::::   :::
 433 nArgLeuArgGluValTyrPheProGlyGlyGluGluValLeuSerThrA 450

2263 .....................................AGTCGCTACACG 2274
                                          ::::::
 450 laAlaLysLysArgArgLysAlaAsnAspAlaAlaAsnLysAsnValTyr 466

2275 GCATTCAGTCTGCAGGATCCATTTGCCCCTACCCAGCATTTAGTTTTAAA 2324
      |||   :::|||   :::      |||   |||   :::
 467 AlaSerAsnLeuAlaGluThrAspAlaGlyThrAlaAsnLysSerSerPr 483

2325 TCTTTTAGATGCTTGGAGAGTTCCCTTAAAAGTGTCTGCAGAGTTGTTGA 2374
      |||                ::::                :::|||:::   :
 483 oLeuGlyGlyLysThrAlaIleLysProProArgGlyAlaGlnAsnLysG 500

2375 TGAGTATAAGACTACCTCCCAAACTG.................AAACAG 2406
      :::::|||:::   |||   |||                    |||:::
 500 lnAsnIleGlnThrProAlaLysThrThrArgThrThrArgGlyLysGlu 516

2407 AATGAAGTAGCTAATGTCCAGCCTTCTTCCCAAAGA.............. 2442
      |||:::::::   :::|||:::   |||:::::|||
 517 AsnAspThrProSerValMetSerSerGlySerArgSerAlaLysAspLy 533

2443 .......GCCAGAATTGAAGACGTACCACCT.................C 2467
            ::::::::::::::::::   |||
 533 sAlaLeuSerLysLeuHisGlnLeuAlaProAspIleAlaLeuTyrGluL 550

2468 CCACTAAAAAGCTAACTCCAGAATTGACCCCTTTTGTGCTTTTCACTGGA 2517
        ||||||   |||   :::      |||:::            :::
 550 ysGluLysLysArgThrAlaLysAspGlyProTrpGlyGlyLysArgAla 566

2518 TTCGAGCCTGTCCAGGTTCAACAGTATATTAAGAAGCTCTACATTCTT.. 2565
      :::   ::::::   ::::::      |||:::
 567 AlaAspGlnIleAspArgAspArgAlaAlaLysGluSerSerProAlaGl 583

2566 ....GGTGGAGAGGTTGCGGAGTCTGCACAGAAG.............. 2595
        |||||||||      |||:::   ||||||
 583 uAspGlyGlyGluGluAspGluThrGluGlnLysArgProAlaLysLysA 600

2595 ..................................................  2595

 600 laArgValSerSerSerProGluMetArgIleCysLeuThrGlyTyrLys 616

2595 ..................................................  2595

 617 ArgTrpValAsnAspLysMetArgGluGluAlaAspArgIleValGlnAs 633

2596 ..........TGCACACACCTCATTGCCAGCAAAGTGACTCGCACCGTGA 2635
                |||   :::|||   |||   :::::::::|||||:::|
 633 pAsnValProCysAspTyrLeuAlaAlaProArgMetValArgThrMetL 650

2636 AGTTCCTGACGGCGATTTCTGTCGTGAAGCACATAGTGACGCCAGAGTGG 2685
      ||||||||   :::::::::            |||::::::   ::::::
 650 ysPheLeuArgCysLeuAlaArgGlyProAspIleIleSerSerAspTyr 666

2686 CTGGAAGAATGCTTCAGGTGTCAGAAGTTCATTGATGAGCAGAACTACAT 2735
      :::      |||   ::::::   |||   :::      ::::::|||::
 667 ValThrAlaCysValGluAlaGlyLysValLeuProProLysAspTyrLe 683

2736 TCTCCGAGATGCTGAGGCAGAAGTACTTTTCTCTTTCAGCTTGGAAGAAT 2785
      :|||   |||   |||:::|||      |||:::   :::|||:::   :
 683 uLeuValAspLysGluSerGluAspArgPheGlyValThrLeuGlnThrA 700

2786 CCTTAAAACGGGCACACGTTTCTCCACTCTTTAAGGCAAAATATTTTTAC 2835
      ::::::::|||||||:::::::::         :::::::::   ::::::
 700 laIleSerArgAlaArgAlaAsn.........ArgGlyArgLeuLeuTrp 713

2836 ATCACACCTGGAATCTGCCCA..............AGTCTTTCCACTAT 2870
        :::|||   :::|||                 ::::::::::::
 714 GlyValProValPheCysThrGluGluIleLysAsnGlyValGlnSerTy 730

2871 GAAGGCAATCGTAGAGTGTGCAGGA 2895
      ::::::|||:::|||:::   |||
 730 rGlnThrIleAlaGluAlaAsnGly 738
```

seq_name: pir1:RGBYS5

seq_documentation_block:
regulatory protein SNF5 - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YBR2036; protein YBR289w
C;Species: Saccharomyces cerevisiae
C;Date: 30-Sep-1991 #sequence_revision 09-Sep-1994 #text_change 21-Jul-2000
C;Accession: S44551; S46171; A36375; S12067; S39145
R;Holmstrom, K.; Brandt, T.; Kallesoe, T.
Yeast 10(Suppl.A), S47-S62, 1994
A;Title: The sequence of a 32420 bp segment located on the right arm of chromosome I
A;Reference number: S44537; MUID:94378722
A;Accession: S44551
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-905 <HOL>
A;Cross-references: EMBL:X76053; NID:g600025; PIDN:CAA53652.1; PID:g429134
R;Brandt, T.; Christiansen, C.; Holmstroem, K.; Kallesoe, T.
submitted to the Protein Sequence Database, August 1994
A;Reference number: S46157
A;Accession: S46171
A;Molecule type: DNA
A;Residues: 1-905 <BRA>
A;Cross-references: EMBL:Z36158; NID:g536741; PIDN:CAA85254.1; PID:g536742; GSPDB:GN
R;Laurent, B.C.; Treitel, M.A.; Carlson, M.
Mol. Cell. Biol. 10, 5616-5625, 1990
A;Title: The SNF5 protein of Saccharomyces cerevisiae is a glutamine- and proline-ri
A;Reference number: A36375; MUID:91042489
A;Accession: A36375
A;Molecule type: DNA
A;Residues: 1-563,'D',565-905 <LAU>
A;Cross-references: GB:M36482; NID:g172637; PIDN:AAA35062.1; PID:g172638
C;Genetics:
A;Gene: SGD:SNF5; MIPS:YBR289w
A;Cross-references: SGD:S0000493; MIPS:YBR289w
A;Map position: 2R
C;Superfamily: regulatory protein SNF5
C;Keywords: nucleus; transcription regulation
F;31-324/Region: glutamine/proline-rich
F;435-683/Region: acidic
F;714-882/Region: proline-rich

alignment_scores:

```
          Quality:  223.50          Length:    303
            Ratio:  1.380             Gaps:      9
Percent Similarity:  53.465   Percent Identity: 27.723

alignment_block:
US-09-664-641-10 x RGBYS5   ..

Align seg 1/1  to: RGBYS5  from: 1  to: 905

     817 CCTCGTCTGATTATTTATGAAGAGGAAGAAGAGGAAGAGGAAGAGGAGGA 866
         |||:::   :::         ::::::::::::::::::      :::::
      42 ProGlnGlnLeuGlnMetIleGlnGlnArgHisGlnGlnLeuLeuArgSe 58

     867 GGAAGTAGAAAATGAGGAACAAGATTCTCAGAATGAGGGTAGTACAGATG 916
         ::::::::::::::::::|||::::::|||                  :
      58 rArgLeuGlnGlnGlnGlnGlnGlnGlnGln.................G 69

     917 AGAAGTCAAGCCCTGCCAGCTCTCAAGAAGGGTCTCCTTCAGGTGACCAG 966
         ::   |||  |||   ::::::::::::::   |||         |||
      69 lnThrSerProProProGlnThrHisGlnSerProProProProProGln 85

     967 CAGTTTTCACCTAAATCCAACACTGAAAAATCTAAAGGGGAATTAATGTT 1016
         |||   :::|||   :::|||   :::
      86 GlnSerGlnProIleAlaAsnGlnSerAla..................... 95

    1017 TCATGATTCTTCAGATTCATCACCGGAAAAACAGGAGAGAAATTTAAACT 1066
            :::|||         |||              :::|||||||:::
      96 .......ThrSerThrProProProProProProAlaProHisAsnLeuHis. 109

    1067 GGACCCCGGCCGAAGTCCCACAGTTAGCTGCAGCAAAACGCAGGCTGCCT 1116
                      |||||||:::
     110 ................ProGlnIle..................... 112

    1117 CAGGGAAAGGAGCCTGGGTTGATTAACTTGTGTGCCAATGTCCCACCCGT 1166
            |||:::   |||                        :::   |||::
     113 ...GlyGlnValPro.....................LeuAlaProAl 120

    1167 CCCAGGTAACATTTTGCCCCCTGAGGTCCGGGGTAATTTAATGGCTGCTG 1216
         :|||   |||   ||||||||||::::::          :::|||:::
     120 aProIleAsn...LeuProProGlnIleAlaGlnLeuProLeuAlaThrG 136

    1217 CACAAAACCTCCAAAGTTCTGAAAGATCAGAAATGATAGCTACCTGGAGT 1266
           |||:::::::   ::::::   |||:::::::   ||||||      :::
     136 lnGlnGlnValLeuAsnLysLeuArgGlnGlnAlaIleAlaLysAsnAsn 152

    1267 CCAGCTGTACGGACACTGAGGAATATTACTAATAATGCTGACATTCAGCA 1316
         |||   |||   :::      ::::::::::::           |||||
     153 ProGlnValValAsnAlaIleThrValAlaGln............GlnGl 165

    1317 GATGAACCGGCCATCAAATGTAGCACATATCTTACAGACTCTTTCAGCAC 1366
         |:::::|||      :::                |||||   ::::::
     165 nValGlnArgGlnIleGluGlnGlnLysGlyGlnGlnThrAlaGlnThrG 182

    1367 CTACGAAAAATTTAGAACAGCAGGTGAATCACAGCCAGCAGGGACATACA 1416
            ::::::   :::|||   :::   ::::::||||||   :::
     182 lnLeuGluGlnGlnArgGlnLeuLeuValGlnGlnGlnGlnGlnGlnGln 198

    1417 AATGCCAATGCAGTGCTGTTTAGCCAAGTGAAAGTGACTCCAGAGACACA 1466
            |||   :::      :::|||   :::         :::
     199 LeuArgAsnGlnIleGlnArgGlnGlnGlnGlnGlnPheArgHisHisVa 215

    1467 CATGCTACAGCAGCAGCAGCAGGCCCAGCAGCAGCAGCAGCAGCACCCGG 1516
         ::::::||||||||||||||   ||||||||||||||||||||:::
     215 lGlnIleGlnGlnGlnGlnGlnLysGlnGlnGlnGlnGlnGlnGln.... 230

    1517 TTTTACACCTTCAGCCCCAGCAGATAATGCAGCTCCAGCAGCAGCAGCAG 1566
            |||   |||   ||||||   :::|||   ||||||||||||||||
     231 .....HisGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 245

    1567 CAGCAGATCTCTCAGCAACCTTACCCCCAGCAGCCGCCGCATCCATTTTC 1616
         ||||||   :::||||||         ||||||      :::      ::
     246 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGl 262

    1617 ACAGCAGCAGCAGCAGCAACAGCAGCCACCACCATCG.........CCTC 1657
         :|||||||||||||||||   |||   |||   |||         ||||
     262 nGlnGlnGlnGlnGlnGlnGlyGlnIleProGlnSerGlnGlnValProG 279

    1658 AGCAGCATCAGCTTTTTGGACATGATCCAGCAGTGGAGATTCCAGAAGAA 1707
         ||   :::::::::   |||:::   |||:::         |||
     279 lnValArgSerMetSerGlyGlnProProThrAsnValGlnProThrIle 295

    1708 GGCTTCTTA 1716
         |||   |||
     296 GlyGlnLeu 298

seq_name: pir2:S69205

seq_documentation_block:
stripe a/b protein - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 21-Jul-2000
C;Accession: S69205
R;Frommer, G.; Vorbrueggen, G.; Pasca, G.; Jaeckle, H.; Volk, T.
EMBO J. 15, 1642-1649, 1996
A;Title: Epidermal egr-like zinc finger protein of Drosophila participates in myotube
A;Reference number: S69205; MUID:96203082
A;Accession: S69205
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-1180 <FRO>
A;Cross-references: EMBL:U42403; NID:g1147788; PIDN:AAB02355.1; PID:g1147789
C;Keywords: alternative splicing


alignment_scores:
          Quality:  220.50          Length:    650
            Ratio:  0.817             Gaps:     26
Percent Similarity:  41.538   Percent Identity: 21.538

alignment_block:
US-09-664-641-10 x S69205   ..

Align seg 1/1  to: S69205  from: 1  to: 1180

     865 GAGGAAGTAGAAAATGAGGAACAAGATTCTCAGAATGAGGGTAGTACAGA 914
         |||:::||||||   :::::::|||           |||:::
     443 GluGlnValGluAlaHisLysGln......LeuIleGluAlaLeuProGl 457

     915 TGAGAAGTCAAGCCCTGCCAGCTCTCAAGAAGGGTCTCCTTCAGGTGACC 964
           :::   :::::::|||:::::::::::::::      ||||||
     457 yAspLeuAsnThrProValThrThrSerSerAspIleProSer..... 471

     965 AGCAGTTTTCACCTAAATCCAACACTGAAAAATCT......AAAGGGGAA 1008
              |||:::|||   :::   :::|||                |||:::
     472 ..PhePheGlyProThrThrValValGluProProProIleThrGlySer 487

    1009 TTAATGTTTGATGAT................................. 1023
         :::      :::|||
     488 IleGluSerGluAspLeuSerLeuGluProGlnValIleSerValAlaSe 504

    1024 .......TCTTCAGATTCATCACCGGAAAAACAGGAGAGAAATTTAAACT 1066
                 |||      ||||||   |||:::|||||||:::
     504 rProValLeuSerHisCysSerProLeuLysGluGluArgSer....... 518

    1067 GGACCCCGGCCGAAGTCCCACAGTTAGCTGCAGCAAAACGCAGGCTGCCT 1116
           ||||||          |||   ||||||   :::
     519 ..ThrPro.........ProAlaLeuAlaIleVal.............. 526

    1117 CAGGGAAAGGAGCCTGGGTTGATTAACTTGTGTGCCAATGTCCCACCCGT 1166
                |||||   :::      |||   |||
     527 ......LysGluGluSerSerAsnAsnSerCys.................. 535

    1167 CCCAGGTAACATTTTGCCCCCTGAGGTCCGGGGTAATTTAATGGCTGCTG 1216
                |||:::   |||   :::   ::::::|||
     536 .......AsnMetTyrProGlnHisAsnAsnAsnAsn.............. 545
```

```
1217 GACAAAACCTCCAAAGTTCTGAAAGATCAGAAATGATAGCTACCTGGAGT 1266
                           :::::::::      :::|||   |||
 546 .......................AsnAsnAsnAsnAsnThrThrSerSer 554

1267 CCAGCTGTACGGACACTGAGGAATATTACTAATAATGCTGACATTCAGCA 1316
     ::::::   |||   ::::::   |||:::::::::::::   :::
 555 SerThrThrThrThrSerLysGlnThrThrSerGluSerAsnThrGluCy 571

1317 GATGAACCGGCCATCAAATGTAGCACATATCTTACAGACTCTTTCAGCAC 1366
      ::::::   |||:::|||   ::::::      |||
 571 sValGlySerProGlyAsnHisThrGlnSerHisGln.............. 583

1367 CTACGAAAAATTTAGAACAGCAGGTGAATCACAGCCAGCAGGGACATACA 1416
           :::   :::||||||:::::|||::::::   :::   :::
 584 ........GlnGlnGlnGlnGlnLeuGlnHisAsnAsnThrSerSerSer 597

1417 AATGCCAATGCAGTGCTGTTTAGCCAAGTGAAAGTGACTCCAGAGACACA 1466
     |||:::|||:::                                  ||
 598 AsnSerAsnCys................................HisHi 603

1467 CATGCTACAGCAGCAGCAGCAGGCCCAGCAGCAGCAG.............. 1503
     |      ||||||||||||||||   ||||||||||
 603 sSerHisGlnGlnGlnGlnGlnGlnGlnGlnGlnHisMetSerSerP 620

1504 ..............CAGCAGCACCCGGTTTTACACCTTCAGCCCCAG... 1536
                   |||||||||   :::||||||   |||   |||
 620 roGlnGlnGlnTyrGlnGlnHisGlnIleLeuHisGlnGlnGlnGlnPhe 636

1537 ..............................................CAGAT 1541
                                                   |||::
 637 GlyTyrHisHisHisHisHisHisHisHisHisHisHisSerGlnLe 653

1542 AATGCAGCTCCAGCAGCAGCAGCAGCAGCAGATCTCTCAGCAACCTTACC 1591
     ::::|||||||||||||||||||||||||||||   :::||||||   :::
 653 uGlnGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnHisG 670

1592 CCCAGCAGCCGCCGCATCCATTTTCACAGCAGCAGCAGCAGCAACAGCAG 1641
       |||||||||   |||      :::|||   ||||:::|||||||||||||
 670 lnGlnGlnProLeuHisGlnGlnGlnGlnLeuGlnHisGlnGlnGlnGln 686

1642 CCACCACCATCGCCTCAGCAGCATCAGCTTTTTGGACATGATCCAGCAGT 1691
                  |||:::||||||          ::::::   |||::
 687 ..............GlnHisHisGlnGlnHisTyrGlnGlnHisAlaIl 698

1692 GGAGATTCCAGAAGAAGGCTTCTTATTGGGATGTGTGTTTGCAATTGCGG 1741
     ::::
 698 eGln.......................................... 699

1742 ATTATCCAGAGCAGATGTCTGATAAGCAACTG...CTGGCCACCTGGAAA 1788
           :::|||:::|||::::::|||            ::::::::
 700 ........HisGlnGlnSerGlnGlnGlnAlaSerLysIleSerTyrArg 713

1789 AGGATAATCCAGGCACATGGCGGCACTGTTGACCCCACCTTCACGAGTCG 1838
       |||:::   :::   |||:::::::::::::   :::   ::::::
 714 GlyIlePheThrThrThrGlyAsnAlaMetAsnAlaAlaAlaAlaAlaAl 730

1839 ATGCACGCACCTTCTCTGTGAGAGTCAAGTCAGCAGCGCGTATGCACAGG 1888
     ::::::            :::   |||   ::::::   :::   |||
 730 aAlaAlaAlaAlaGlnGlnGlnHisGlnGlnGlnGlnGlnHisGlnGlnG 747

1889 CAATAAGAGAAAGAAAGAGATGTGTTACTGCACACTGGTTAAACACAGTC 1938
       :::   :::   :::      |||   |||
 747 lnLeuProSerProGlnLeuGlyValLeuAla.................. 757

1939 TTAAAAAAAAAAAAAATGGTACCGCCGCACCGAGCCCTTCACTTCCCAGT 1988
                    |||   ||||||   ::::::|||
 758 .........GlyProMetSerProProSerAsnSerLeuGlyAsnSerTr 771

1989 GGCCTTCCCACCAGGAGGAAAGCCATGTTCACAG.................. 2022
     :::::::|||          |||          |||
 771 pGlyLeuProSerProAspLysThrMetPheGlnProProLeuPheSerL 788

2023 ........CATATTATTTCTGTG........................... 2037
```

```
                |||      ::::::
 788 euProAlaHisTyrAlaThrMetGlnGlnGlnGlnGlnGlnGlnGlnGln 804

2038 ..............................ACTGGATTTGTTGATAGTGA 2057
                                    :::|||   :::   |||
 805 GlnGlnGlnGlnGlnGlnGlnGlnGlnAlaAlaGlyAlaAlaProSerPr 821

2058 CAGAGATGACCTAAAATTAATGGCTTATTTGGCAGGT......GCCAAAT 2101
       ||||||   :::      |||      |||:::      |||:::
 821 oTyrAspAspGlyArgAlaAlaAlaAlaAlaAlaAlaGlnHisAlaGluL 838

2102 ATACGGGTTATCTATGCCGCAGCAACACAGTCCTCATCTGTAAAGAACCA 2151
       |||   |||      :::   |||   |||:::   |||:::|||
 838 euLeuGly...LeuThrMetAspCysThrProLeuLeuLeuLysGlnPro 853

2152 ACTGGTTTAAAGTATGAAAAAGCCAAAGAGTGGAGGATACCCTGTGTCAA 2201
              :::|||      |||:::                  :::::
 854 ProPro...SerTyrAlaGlyAlaSer.................AlaGl 863

2202 CGCCCAGTGGCTTGGCGACATTCTTCTGGGAAACTTTGAGGCACTGAGGC 2251
     :         ||||||||||:::      ::::::   :::   |||:::|
 863 yPheProGlyLeuGlyAspLeuHisSerSerHisGluGlnGlnLeuGlnG 880

2252 AGATTCAGTATAGTCGCTACACGGCATTCAGTCTGCAGGATCCATTTGCC 2301
     ||   ||||||
 880 lnGlnGlnTyr..................................... 883

2302 CCTACCCAGCATTTAGTTTTAAATCTTTTAGATGCTTGGAGAGTTCCCTT 2351

 883 ..................................................... 883

2352 AAAAGTGTCTGCAGAGTTGTTGATGAGTATAAGACTACCTCCCAAACTG. 2400
                                    :::|||      ||||||
 884 ........................ValArgSerGlnProLysTyrG 891

2401 ............................AAACAGAATGAAGTAGCT 2418
                                   :::|||:::::::|||
 891 lnTrpLeuAspSerProAlaAspTyrAlaGlnGlnGlnGlnGlnValGln 907

2419 AATGTCCAGCCTTCTTCCCAAAGAGCCAGAATTGAAGACGTACCACCTCC 2468
     :::||||||   ::::::|||:::   :::          :::|||   ||
 908 GlnValGlnGlnGlnGlnGlnGlnGlnGlnThrLeuValLeuProGlyPr 924

2469 CACTAAAAAGCTAACT................................ 2484
     ||||::::::   :::
 924 oThrSerSerAlaSerSerSerAsnAlaAlaLeuGlyValLeuIleProL 941

2485 ..............CCAGAATTGACCCCTTTTGTGCTTTTCACTGGATTC 2520
                    |||::::::   |||            |||||:::
 941 ysGlnGluAsnTyrProAspMetGlnProSerSerAsnGlyThrGlyTyr 957
```

seq_name: pir2:T14577

seq_documentation_block:
protein kinase YakA (EC 2.7.1.-) - slime mold (Dictyostelium discoideum)
C;Species: Dictyostelium discoideum
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T14577
R;Kuspa, A.; Lu, S.; Souza, G.M.
submitted to the EMBL Data Library, January 1998
A;Description: YakA, a protein kinase required for the growth to development transit
A;Reference number: Z18146
A;Accession: T14577
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1457 <KUS>
A;Cross-references: EMBL:AF045453; NID:g2854116; PID:g2854117; PIDN:AAC02554.1
C;Genetics:
A;Gene: yakA
C;Keywords: ATP; phosphoprotein; phosphotransferase; serine/threonine-specific prote

alignment_scores:

```
          Quality:  219.50          Length:    397
            Ratio:  1.155             Gaps:     15
Percent Similarity:  47.859  Percent Identity: 23.929

alignment_block:
US-09-664-641-10 x T14577   ..

Align seg 1/1  to: T14577  from: 1  to: 1457

  835 GAAGAGGAAGAAGAGGAAGAGGAAGAGGAGGAGGAAGTAGAAAATGAGGA 884
      :::::::::::::::::::::::::::::::::::   :::|||:::
  614 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnHisAsnGlnPh 630

  885 ACAAGATTCTCAGAATGAGGGTAGTACAGATGAGAAGTCAAGC..... 927
       |||:::::::|||:::::::   :::   ::::::::::|||||
  630 eGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnSerSerSerGlnG 647

  928 .....CCTGCCAGCTCT 939
                     |||   ::::||
  647 lnHisIleGlnIleGlnProLeuGlnLeuPheSerThrProTyrThrSer 663

  940 CAAGAAGGGTCTCCT..................TCAGGT........... 960
      ::::::::::|||                    ||||||
  664 ThrAsnAsnThrProAsnLeuSerSerSerAsnSerGlySerSerLeuAs 680

  961 .....GACCAGC 967
                 :::||||
  680 nAsnLeuLysLysLeuAsnLeuProProPheLysGlnGlnGlnGlnGlnG 697

  968 ACTTTTCACCTAAATCCAACACT.....................GAAAAA 996
      ||||||||   ::::::|||:::                     |||:::
  697 lnPheSerThrSerGlnAsnSerAspSerPheAsnPheProAsnGluSer 713

  997 TCTAAAGGGGAATTAATGTTTGATGATTCTTCAGATTCATCACCGGAAAA 1046
         :::   :::   ::::::      ||||||   |||:::   :::::
  714 PheSerProArgGlyIleTyrIleProSerSerAlaSerAsnIleGlnGl 730

 1047 ACAGGAGAGA.................AATTTAAACTGGACCCCGGCCG 1078
      :|||::::::                 |||   |||   :::   ::::
  730 nGlnGlnGlnProIleAsnIleAsnAsnAsnGlnAsnGlyValGlySerG 747

 1079 AAGTCCCACAGTTAGCTGCAGCAAAACGCAGGCTGCCTCAGGGAAAGGAG 1128
      ::|||   |||||||||                          |||::::::
  747 lnValSerGlnLeuAlaLeu.....................GlyGlnSer 756

 1129 CCTGGGTTGATTAACTTGTGTGCCAATGTCCCACCCGTCCCAGGTAACAT 1178
      |||:::|||::::::                      :::|||   |||||
  757 ProSerLeuPheGly.....................ThrProThrAsnIl 766

 1179 TTTGCCCCCT......GAGGTCCGGGGTAATTTAATGGCTGCTGGACAAA 1222
      |   ||||||      ::::::   :::|||      :::|||      |
  766 eTyrProProTyrSerSerMetTyrAsnAsnSerProValAlaThrProA 783

 1223 ACCTCCAAAGT....................TCTGAAAGATCAGAAATG 1251
      ||      |||                    |||:::   |||::::::
  783 snSerLeuSerPheTyrGlySerSerTrpGlySerAspSerSerSerIle 799

 1252 ATAGCTACCTGGAGTCCAGCTGTACGGACACTGAGGAATATT........ 1293
         :::   :::|||::::::   |||   :::::::::
  800 ......SerLeuAsnProSerThrProThrGlnLysGlnMetPheGlnGl 814

 1294 ...............ACTAATAATGCTGACATTCAGCAGATGAACCGGC 1327
                      :::||||||   :::   ::::::   |||:::
  814 nGlnGlnTyrSerAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnA 831

 1328 CATCAAATGTAGCACATATCTTACAGACTCTTTCAGCACCTACGAAAAAT 1377
        ||||||      :::      :::|||   :::      :::::|||
  831 snSerAsnAsnAsnAsnGlyAsnAsnThrAsnAsnIleAsnSerAsnAsn 847

 1378 TTAGAACAGCAGGTGAATCACAGCCAGCAGGGACATACAAATGCCAATGC 1427
         ::::::::::||||||:::   :::::::::   ::::::      ::
  848 AsnAsnAsnAsnValAsnArgArgAsnArgSerLysSerAspIleProSe 864

 1428 AGTGCTGTTTAGCCAAGTGAAA...GTGACTCCAGAGACACACATGCTAC 1474
      :      |||||||:::   :::   :::   |||:::   ::::::   |
  864 rAspSerPheSerSerSerGluGlyMetAspProGlnPheAsnLeuTyrG 881

 1475 AGCAGCAGCAGCAGGCCCAGCAGCAGCAGCAGCACCCGGTTTTACAC 1524
      ||||||||||||||   ||||||||||||||||||:::      :::
  881 lnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln......Gln 895

 1525 CTTCAGCCCCAGCAGATAATGCAGCTCCAGCAGCAGCAGCAGCAGAT 1574
         |||   ||||||   :::|||   ||||||||||||||||||||
  896 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGl 912

 1575 CTCTCAGCAACCTTACCCCCAGCAG..................... 1599
       :::|||      |||   ||||||
  912 nGlnGlnLeuGlnTyrGlnGlnGlnPheGlnThrLeuGlnAspLeuAsnI 929

 1600 .................CCGCCGCATCCATTTTCA.............. 1617
                      |||   :::|||   |||
  929 leGluGlyGluLysProProIleTyrProAsnSerProHisArgLysArg 945

 1617 ..... 1617

  946 SerHisSerGlyTyrLeuAspGlnTyrAlaAsnGlyTyrAsnSerGlnGl 962

 1618 .............................CAGCAGCAGCAGCAGC 1633
                                  ||||||||||||||||
  962 nThrAsnAsnThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnG 979

 1634 AACAGCAGCCACCACCATCGCCTCAGCAGCATCAGCTTTTT 1674
      ||||||||         :::   |||||||||      |||
  979 lnGlnGlnGlnGlnGlnGlnGlnGlnGlnHisProSerPhe 992

seq_name: pir2:S37876

seq_documentation_block:
Glutamine-rich protein YKL054c - yeast (Saccharomyces cerevisiae)
C;Alternate names: protein YKL308
C;Species: Saccharomyces cerevisiae
A;Variety: strain S288C
C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 29-Oct-1999
C;Accession: S37876; S39184; S44529
R;Rasmussen, S.; von Wettstein, D.
submitted to the Protein Sequence Database, March 1994
A;Reference number: S37872
A;Accession: S37876
A;Molecule type: DNA
A;Residues: 1-738 <RAS>
A;Cross-references: EMBL:Z28054; NID:g486148; PIDN:CAA81890.1; PID:g486149; MIPS:YKLC
R;Rasmussen, S.W.
submitted to the EMBL Data Library, November 1993
A;Reference number: S39168
A;Accession: S39184
A;Molecule type: DNA
A;Residues: 1-738 <RA2>
A;Cross-references: EMBL:X75781; NID:g433634; PIDN:CAA53418.1; PID:g433643
R;Rasmussen, S.W.
Yeast 10, 63-68, 1994
A;Title: Sequence of a 28.6 kb region of yeast chromosome XI includes the FBA1 and TC
otifs also found in plant storage proteins and 13 ORFs with weak or no homology to kn
A;Reference number: S44521
A;Accession: S44529
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 386-504 <RAW>
A;Cross-references: EMBL:X75781
C;Genetics:
A;Map position: 11L


alignment_scores:
          Quality:  213.00          Length:    324
            Ratio:  1.275             Gaps:     11
Percent Similarity:  51.543  Percent Identity: 26.543
```

```
alignment_block:
 US-09-664-641-10 x S37876   ..

 Align seq 1/1  to: S37876  from: 1  to: 738

     835 GAAGAGGAAGAAGAGGAAGAGGAAGAGGACGACGAAGTAGAAAATGAGGA 884
         |||:::|||:::|||||||:::::::::::::|||      :::|||:::|| 
     199 GluLysGluGlnGluGluGlnGlnGlnGlnGluGlyHisAsnAsnLysGl 215

     885 ACAAGATTCTCAGAATGAG............................... 903
         |:::   :::|||   |||
     215 uGluHisLysGlnIleGluGlnProSerLeuSerSerLysLysThrThrS 232

     904 ..........GGTAGTACAGATGAGAAGTCAAGCCCTGCCAGCTCTCAA 942
                 :::|||      :::|||   |||   |||:::
     232 erArgThrSerAlaSerGlnProLysLysMetSerTrpAlaAlaIleAla 248

     943 GAAGGGTCTCCTTCAGGTGACCAGCAGTTT...TCACCTAAATCCAACAC 989
           :::|||::::::   ::::::      ||||||   :::|||::
     249 ThrProLysProLysAlaValLysLysThrGluSerProLeuGluAsnVa 265

     990 TGAAAAATCTAAAGGGGAATTAATGTTTGATGATTCTTCAGATTCATCAC 1039
         :   :::   |||   |||:::              ::::::|||:::::: 
     265 lAlaGluLeuLysLysGluIleSerAspIleLysLysAspAspGlnLysS 282

    1040 CGGAAAAACAGGAGAGAAATTTAAACTGGACCCCGGCCGAAGTCCCACAG 1089
           |||   :::|||:::::::::|||                      ||| 
     282 erGluAlaSerGluGluLysValAsn.................GluGln 292

    1090 TTAGCTGCAGCAAAACGCAGGCTCCCTCAGGGAAAGGAGCCTGGGTTGAT 1139
         ::::::|||::::::::::      :::      |||||||:::
     293 GluThrSerAlaGlnGluGlnGluGluGluThrAlaGluProSerGluGl 309

    1140 TAACTTGTGTGCCAATGTCCCACCCGTCCCAGGTAACATTTTGCCCCCTG 1189
           |||          :::||||||   |||   |||:::            |
     309 uAsnGlu...AspArgValProGluValAspGlyGlu.............G 321

    1190 AGGTCCGGGGTAATTTAATGGCTGCTGGACAAAACCTCCAAAGTTCTGAA 1239
         |||||:::                                 ::::::::::|||
     321 luValGln.............................GluGluAlaGlu 327

    1240 AGATCAGAAATGATAGCTACCTGGAGTCCAGCTGTACGGACACTGAGGAA 1289
         :::::::|||::::::          :::   :::::::::   |||:::::
     328 LysLysGluGlnValLysGluGluGluGlnThrAlaGluGluLeuGluGl 344

    1290 TATTACTAATAAT......GCTGACATTCAGCAGATGAACCGGCCATCAA 1333
         :      :::|||      |||      ::::::::::::      ::::
     344 nGluGlnAspAsnValAlaAlaProGluGluGluValThrValValGluG 361

    1334 ATGTAGCACATATCTTACAGACTCTTTCAGCACCTACGAAAAATTTAGAA 1383
         ::   :::::::|||      :::::::|||   |||   ::::::
     361 luLysValGluIleSerAlaValIleSerGluProProGluAspGlnAla 377

    1384 CAGCAGGTGAATCACAGCCAGCAGGGACATACAAATGCCAATGCAGTGCT 1433
         :::   |||   :::   ||||||
     378 AsnThrValProGlnProGlnGln.......................... 385

    1434 GTTTAGCCAAGTGAAAGTGACTCCAGAGACACACATGCTACAGCAGCAGC 1483
               |||   :::         :::   ::::::   ||||||   |
     386 .......GlnSerGlnGlnProGlnGlnProGlnGlnProGlnGlnProG 400

    1484 AGCAGGCCCAGCAG...CAGCAGCAGCAGCACCCGGTTTTACACCTTCAG 1530
         |||||   ||||||   ||||||||||||:::|||              |||
     400 lnGlnProGlnGlnProGlnGlnGlnGlnGlnProGlnGlnProGlnGln 416

    1531 CCCCAGCAGATAATGCAGCTCCAGCAGCAGCAGCAGCAG........... 1569
         |||||||||   :::|||   ||||||||||||||||||
     417 ProGlnGlnGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnProValGl 433

    1570 .......................CAGATCTCTCAGCAACCTTACCCCC 1594
                                 |||:::|||||||:::   |||   |
     433 nAlaGlnAlaGlnAlaGlnGluGluGlnLeuSerGlnAsnTyrTyrThrG 450

    1595 AGCAGCCGCCGCATCCATTTTCACAGCAG..................... 1623
         |||||      :::   ::::::||||||
     450 lnGlnGlnGlnGlnGlnTyrAlaGlnGlnGlnHisGlnLeuGlnGlnGln 466

    1624 .........................CAGCAGCAGCAACAGCAGCCACC  1646
                                  ||||||||||||||||:::|||
     467 TyrLeuSerGlnGlnGlnGlnTyrAlaGlnGlnGlnGlnGlnHisProGl 483

    1647 ACCATCGCCTCAGCAGCATCAG 1668
         |||:::   ||||||   |||
     483 nProGlnSerGlnGlnProGln 490

seq_name: pir2:T13167

seq_documentation_block:
Lola-like protein - fruit fly (Drosophila hydei)
C;Species: Drosophila hydei
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C;Accession: T13167
R;Cavarec, L.; Jensen, S.; Casella, J.F.; Cristescu, S.A.; Heidmann, T.
Mol. Cell. Biol. 17, 482-494, 1997
A;Title: Molecular cloning and characterization of a transcription factor for the co
A;Reference number: Z17624; MUID:97127405
A;Accession: T13167
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1010 <CAV>
A;Cross-references: EMBL:Y14994; NID:g2467111; PIDN:CAA75228.1; PID:g2467112
C;Genetics:
A;Cross-references: FlyBase:FBgn0022628
C;Superfamily: POZ domain homology
F;18-119/Domain: POZ domain homology <POZ>


alignment_scores:
            Quality:  210.50              Length:    334
              Ratio:  1.126                Gaps:     15
 Percent Similarity:  55.988   Percent Identity: 26.347

alignment_block:
 US-09-664-641-10 x T13167   ..

 Align seq 1/1  to: T13167  from: 1  to: 1010

     769 CTGGATTGCGTATCAGAGAAAACCAAA...AAGGACGAAGCATTTTATCA 815
         ||||||      ::::::|||   |||   ||||||:::::::::
     362 LeuAspLysSerAsnHisLysGlnLysIleLysAspAsnSerIleThrTh 378

     816 TCCTCGTCTGATTATTTATGAAGAGGAAGAAGAGGAAGAGGAAGAGGAGG 865
                 :::|||      :::   |||   :::::::::   ::::::|
     378 rGluMetValIleGluProLysAlaGluTyrAspAspAspAlaHisAspG 395

     866 AGGAAGTAGAAAATGAGGAACAAGATTCTCAGAATGAGGGTAGTACAGAT 915
         ||:::||||||:::          |||   :::::::::   :::   :::
     395 luAsnValGluAspLeuThrLeuAsp...GluGluAspMetThrMetGlu 410

     916 GAGAAGTCAAGCCCTGCCAGCTCTCAAGAAGGGTCTCCTTCAGGTGACCA 965
         |||   ::::::   |||:::::::::::::|||:::   ::::::::::||
     411 GluLeuAspGlnAlaAlaGlyThrSerGlnGlyGlyGluGlySerSerGl 427

     966 GCAGTTTTCA......CCTAAATCCAACACTGAAAAATCTAAAGGGGAAT 1009
         |   ::::::          :::::::::::   :::   |||   :
     427 nAlaTyrAlaThrTrpGlnHisAspArgSerGlnAspGluLeuGlyLeuM 444

    1010 TAATGTTTGATGATTCTTCAGATTCATCACCGGAA......AAACAGGAG 1053
         ::      :::|||::::::::::   :::|||:::      ||||||:::
     444 etAlaAlaGlnAspAlaGlnGlnArgAspProGlnAspAlaLysGlnAsp 460

    1054 AGAAATTTAAACTGGACCCCGGCC............GAAGTCCCACAG.. 1089
         ::::::   :::          |||           |||:::   :::
     461 LysGlyGluGlnThrGluGlyAlaHisAspGluPheGluLeuAspAspCy 477
```

```
 1090 ....TTAGCTGCAGCAAAACGCAGGCTGCCTCAGGGAAAGGAGCCTGGGT 1135
       |||   :::           :::   |||:::|||:::
  477 sLeuLeuGluSerAsnAspIleValIleThrGlnAsnLysAsp..... 491

 1136 TGATTAACTTGTGTGCCAATGTCCCACCCGTCCCAGGTAACATTTTGCCC 1185
                                        |||   :::|||
  492 ..............................GlyPheValLeuHis 496

 1186 CCTGAGGTCCGGGGTAATTTAATGGCTGCTGGACAAAACCTCCAAAGTTC 1235
       :::     ||||||:::  ||||||  ::::::     :::::
  497 ValLysLysLeuGlyAsnIleThrAlaAlaLysMetGlySerThrThrAs 513

 1236 TGAAAGATCAGAAATGATAGCTACCTGGAGTCCAGCTGTACGGACACTGA 1285
       :|||::::::     :::::::::  :::||||||  :::     :
  513 pGluGlnAlaAlaAlaValThrValThrGlyProAlaGlyGlnProThrG 530

 1286 GGAATATTACTAATAATGCTGACATTCAGCAGATGAACCGGCCATCAAAT 1335
       :::::||||||:::        ::::::|||     :::  |||:::
  530 lnThrIleThrGluLeu......LeuSerGlnAlaAlaHisSerAsp 544

 1336 GTAGCACATATCTTACAGACTCTT...TCAGCACCTACGAAA..... 1374
       |||        |||  ||||||  :::::::|||  |||
  545 ValLysProThrLeuThrThrLeuThrAsnThrProIleLysLeuProSe 561

 1375 .................AATTTAGAACAGCAGGTG............ 1392
                      |||:::::::::  :::
  561 rSerGluCysGluLeuIleAsnIleLysLysIleIleProSerSerThrT 578

 1393 .................AATCACAGCCAGCAGGGACATACAAATGCC 1422
                       :::  |||::::::|||  |||::::::
  578 hrLeuAlaThrHisGlyGlySerSerSerSerSerGlyGlyThrThrThr 594

 1423 AATGCAGTGCTGTTTAGCCAA.....GTGAAAGTGACTCCAGAGAC 1463
       ::::::::::::    :::                |||||  |||||
  595 ThrThrIleIleHisProHisHisIleIleGlnHisValThrGlnGluTh 611

 1464 A.....CACATGCTACAGCAGCAGCAGGCCCAGCAGCAGCAGC 1504
       |       |||:::  |||||||||||||:::  ||||||||||||
  611 rGlnGlnGlnHisGlnGlnGlnGlnGlnGlnHisGlnGlnGlnGlnG 628

 1505 AGCAGCACCCGGTTTTACACCTTCAGCCCCAGCAGATAATGCAGCTCCAG 1554
       |||||:::              |||  |||:::     |||  |||
  628 lnGlnGlnGln...........GlnGlnGlnHisHisHisGlnGlnGln 640

 1555 CAGCAGCAGCAGCAGCAGATCTCTCAGCAACCTTACCCCCAGCAGCCGCC 1604
       |||||||||||||||||:::      :::     ||||||:::
  641 GlnGlnGlnGlnGlnGlnLeuHisIleGluGluValProGlnSerSerHi 657

 1605 GCATCCATTTTCACAGCAGCAGCAGCAGCAACAGCAGCCACCACCATCGC 1654
       |||      :::|||||||||||||||||||||||| ||||||||||:::
  657 sHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGln.HisHisHisGln 673

seq_name: pir2:S44898

seq_documentation_block:
ZK1236.3 protein - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 14-Sep-1994 #sequence_revision 12-May-1995 #text_change 23-Mar-2001
C;Accession: S44898
R;Favello, A.D.
submitted to the EMBL Data Library, May 1993
A;Description: Sequence of the C. elegans cosmid ZK1236.
A;Reference number: S44622
A;Accession: S44898
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1000 <FAV>
A;Cross-references: EMBL:L13200; NID:g289748; PID:g289754
C;Genetics:
A;Introns: 56/3; 86/2; 189/3; 243/2; 669/3; 808/3; 984/3


alignment_scores:

            Quality:  204.50              Length:   1006
              Ratio:  0.492                Gaps:     47
Percent Similarity:  41.352   Percent Identity: 19.682

alignment_block:
US-09-664-641-10 x S44898   ..

Align seg 1/1  to: S44898  from: 1  to: 1000

  139 AGGGCTACGGATCTCTCGGCCCCTGCCCTTCAAATCACCTCTAGGCACAG 188
       |||      |||  |||  |||  |||:::     |||:::
  104 ArgMetLysAspThrSerLeuProProLeuGluAsnProSerAsnGlyGl 120

  189 GTTATTCAGCTTCTCAAGGCTGGAAAAGCGAAGGAAGTTTCCTACAATGC 238
       :::  |||:::  :::                 ||||||:::::
  120 yThrTyrLeuPheAsnLeuIle.........ThrValPheProSerIleA 134

  239 ACTAGCCTCACACATAATCTCAGAGGATGGGGACAATCCAGAGGTGGGAG 288
            :::  |||  |||      |||            :::
  134 laLeuSerPheAsnTyrSerTrpArgGlyAspThrGlyProSerIleSer 150

  289 AAGCTCGGGAAGTCTTTGACTTACCTGTTGTAAAGCCTTCTTGGGTGATT 338
       ::::::  ||||||:::  |||:::  ||||||:::
  151 IlePheSerPheSerLeuSerValLeuPhePheSerLeuPheProGlnHi 167

  339 CTGTCCGTTCAGTGTGGAACTCTTCTGCCAGTAAATGGTTTTTCTCCAGA 388
            :::  ::::::  :::|||
  167 sLysAsnIleCysAlaArgAlaTrpCys..................... 176

  389 ATCATGTCAGATTTTTTTTGGAATCACTGCCTGCCTTTCTCAGGGTGTTG 438
            |||  |||  :::|||||||        ||||||  :::
  177 ........ArgProPheArgArgSerLeuSerPheSerLeuArgPhelle 190

  439 ATACAAGCTGCAGCTCTTTGTTGGAGTCTTCCAGAGCTCTCCCAAGGAGA 488
       ::::::                |||  |||      |||:::  ||
  191 MetSer.................SerGluProAlaSerSerSerThrGl 201

  489 G...GTAGGGAAGGGAGCTTGTCCAGCAGAAGTTGGGAAGCACAGAGATC 535
       |   |||  :::          |||  ::::::  ||||||:::  :
  201 uLysValProGluGluProHisProHisSerIle...LysHisLysPheG 217

  536 ATCTGCCTTCTTCTGACCCGGTATTGATGCAGGCTGAGGCCTCTGTTGTA 585
       ::  |||:::         |||:::  :::|||  ::::::  |||
  217 lnGlyProGlnPhe......ValIleProArgAlaLeuSerAspTyrVal 231

  586 ATGTGCTGGGTGTCATCTGAAGACAGAAGTGCCCTGTGGGCTTTGGTTAC 635
       :::         ::::::      :::|||          ||||||  ::
  232 LeuAsnValAsnAsnGlnThrProGluSerTyrGluLysAlaLeuAsnAl 248

  636 GTTCTATGGGGGAGATTGCCAGCTAACCCTCAATAAGAAATGCACGCATT 685
       :  ||||||  |||      :::||||||          |||  |||:
  248 aLysTyrGlyArgAspAspTyrIleThrLeu.........CysLeuHisV 262

  686 TGATTGTT.................CCAGAGCCAAAGGGGGGGAAATAC 717
       ::                      |||         |||  :::|||
  262 alThrSerLeuCysThrGluTyrGlyProLeuAspAspGlyProGluTyr 278

  718 GAA......TGTGCTTTAAAGCGAGCAAGTATTAAAATTGTG........ 753
            |||      :::          |||::::::
  279 ValLeuL.uCysAspSerArgAlaArgValAspLysLeuIleGluAspLe 295

  753 ..................................................... 753

  295 uValLysLeuGluIleAspThrAspTyrValGlnLeuGluLeuHisG 312

  754 ....................ACTCCTGACTGGGTTCTGGATTGCGTA 780
                           ||||||  ||||||      :::
  312 lyGlyLysArgLeuHisLeuGlnLysProAspAlaValLeuArgAspIle 328

  781 TCAGAGAAAACCAAA.........AAGGACGAAGCATTTTATCATCCTCG 821
       :::  |||  :::         :::|||:::  |||  :::  ::
  329 AlaTyrLysGlnSerAsnGluGlySerAspLysPhePheLeuGluMetLy 345
```

```
 822 TCTGATT.......................ATTTATGAAGAGGAAGAAG 847
     :|||:::                        |||   ::::::|||||||
 345 sLeuValProSerGluSerMetLysAlaLysIleMetLysGlnGluGluG 362

 848 AGGAAGAG.......................GAAGAGGAGGAGGAAGTA 873
     ||||||||                       ::::::   :::|||
 362 luGluGluLysAlaArgLysHisGlyGlnTyrGlnGlnTyrGlnGluTyr 378

 874 GAAAATGAGGAACAA.........GATTCTCAGAATGAGGGTAGTACAGA 914
     :::::::::::::|||         |||:::|||::::::::::|||:::
 379 HisGlnGlnHisGlnAlaMetAsnAspGlyGlnSerSerSerSerValPr 395

 915 TGAGAAGTCAAGCCCTGCCAGCTCTCAAGAAGGGTCT............. 951
      :::   ||||||||||:::   |||:::|||::::::
 395 oSerThrSerSerProSerCysSerSerGluAlaAsnArgLysGluMetG 412

 952 .............CCTTCAGGTGACCAGCAGTTT.............. 972
                  |||:::|||   ::::::::
 412 luThrValArgGluProAlaGlyProSerGluLeuMetArgAlaIleAsn 428

 973 ....................................TCACCTAAATC 983
                                        :::|||   ::
 429 AlaProValAlaProAlaProValValIleLysIleGluThrProValAl 445

 984 CAACACTGAAAAATCTAAAGGGGAATTAATGTTTGATGATTCTTCAGATT 1033
     :     |||:::::::::      ||||||   ||||||:::      |
 445 aLeuProGluGluAspGlu...ThrLeuMetAspAspAspGluMetProS 461

1034 CA..................................................... 1035
     ||
 461 erLeuThrValGluAlaProSerGluGluAlaSerPheGluAlaGluGln 477

1036 ...TCACCGGAAAAACAGGAGAGAAATTTAAACTGG......ACCCCGGC 1076
        |||||||:::      :::   :::::::::
 478 ProSerProGlnValProGlnAlaSerIleGluGlyProSerGlnGlnGl 494

1077 CGAAGTCCCACAGTTAGCTGCAGCAAAACGCAGGCTGCCTCAGGGAAAGG 1126
     :::::::|||      :::      ||||||:::::::   :::|||::::
 494 nGlnIleProGlyThrSerGlnGlnLysArgGlnValAlaArgGlySerA 511

1127 AGCCTGGGTTGATTAACTTGTGTGCCAATGTCCCACCCGTCCCAGGTAAC 1176
     ::   ::::::::|||:::         ::::::||||||      ||||||
 511 rgThrAsnMetIleSerTyr...HisAspLeuProProGlyThrGlyAsn 526

1177 ATTTTGCCCCCTGAGGTCCGGGGTAATTTAATGGCTGCTGGACAAAACCT 1226
           ||||||                 ||||||::   |||
 527 ...AlaProPro.................MetAlaCysProGlnValTh 536

1227 CCAAAGTTCTGAAAGATCAGAAATGATAGCTACCTGGAGTCCAGCTGTAC 1276
        :::   |||::::::          :::   ::::::      :::|
 536 rLeuLysLeuGluLysAsn.........ValProPheGluAlaLysIleA 550

1277 GGACACTGAGGAATATTACTAATAAT......GCTGACATTCAGCAGATG 1320
     ||::::::   :::   |||:::::::     ::::::::::|||:::|||
 550 rgAlaValAlaGlyTyrThrArgLysProIleSerGluValGlnLysMet 566

1321 AACCGGCCATCAAATGTAGCACATATCTTACAGACTCTTTCAGCACCTAC 1370
        |||||||||:::::::      |||:::::::::::::
 567 ...ArgProSerAspLeuGluSerIlePheHisSerIle........... 578

1371 GAAAAATTTAGAACAGCAGGTGAATCACAGCCAGCAGGGACATACAAATG 1420

 578 ..................................................... 578

1421 CCAATGCAGTGCTGTTTAGCCAAGTGAAAGTGACTCCAGAGACACACATG 1470
                ::   |||      ::::::      ::::::::::::
 579 ........CysIleAlaSerValGlnArgIleLysArgArgAsnGluLeu 592

1471 CTACAGCAGCAGCAGCAGGCCCAGCAGCAGCAGCAGCAGCACCCGGTTTT 1520
     :::|||||||   |||:::   :::   |||:::   :::   |||:::::
 593 ValGlnGlnLeuGlnGluIleAsnAlaGlnSerCysLysSerProThrMe 609

1521 ACACCTTCAGCCC.....................CAGCAGATAATG. 1545
```

```
     :   ::::::                              |||:::::::::
 609 tThrMetAsnLysLysPheThrLeuAlaLysAlaTyrGlnArgValGlnA 626

1546 ..CAGCTCCAGCAGCAGCAGCAGCAGATCTCTCAGCAACCTTACCCC 1593
       ::::::::::::   ::::::::::||||||      |||   |||
 626 snGluIleGluLysIleAspArgGluGlnIleLeuProGlnGlnTyrMet 642

1594 CAGCAGCCGCCGCATCCATTTTCACAGCAGCAGCAGCAGCAACAGCAGCC 1643
     :::::::||||||   |||            |||   ||||||:::   ||
 643 AsnMetProProMetPro.........ProGlnGlyGlnGlnArgLeuPr 656

1644 ACCACCATCG.....................CCTCAGCAGCATCAGCTTT 1672
     |||||||:::                          |||||||||::::::
 656 oProProAlaTyrProProGlyIleLeuProProGlnGlnAsnArgGlnG 673

1673 TTGGACATGATCCAGCAGTGGAGATTCCAGAAGAAGGCTTCTTATTGGGA 1722
       |||      |||      :::          :::   |||:::::|||
 673 lnGlyValProProGlnPheGlnArgSerProGln...PheMetIleGly 688

1723 TGTGTGTTTGCAATTGCGGATTATCCAGAGCAGATGTCTGATAAGCAACT 1772
                                           |||   |||
 689 ..................................ProAspGlyGlnAr 693

1773 GCTGGCCACCTGGAAAAGGATAATCCAGGCACATGGCGGCACTGTTGACC 1822
                                        ||||||          |
 693 g.......................TyrAlaHisProTyrMetGlnLeuP 702

1823 CCACCTTCACGAGTCGATGCACGCACCTTCTCTGTGAGAGTCAAGTCAGC 1872
     ||:::   :::::::|||   ::::::::|||      |||:::|||:::
 702 roAsnSerAsnGlnArg...AlaArgIleLeuAsnThrSerSerValGln 717

1873 AGCGCGTATGCACAGGCAATAAGAGAAAGAAAGAGATGTGTTACTGCACA 1922
               ::::::   :::|||:::|||          :::   |||
 718 Pro......SerGluGluValArgAsnArgLeuValLysIleGluAlaMe 732

1923 CTGGTTAAACACAGTCTTA.............AAAAAAAAAAAAATGGTAC 1960
        :::|||   :::              :::              |
 732 tAlaMetAsnMetAlaGlnLeuAsnProProArgProProProProGlnP 749

1961 CGCCGCACCGAGCCCTTCACTTCCCAGTGGCCTTC......CCACCAGGA 2004
     ||||||||||||||||||:::      :::   |||      |||   :::
 749 roProHisArgAlaLeuGlnGlyGluLeuGlnPheLeuArgProGlyAla 765

2005 GGAAAGCCATGTTCACAGCATATTATTTCTGTGACTGGATTTGTTGATAG 2054
           ||||||:::
 766 ProAspProCysAsn..................................... 770

2055 TGACAGAGATGACCTAAAATTAATGGCTTATTTGGCAGGTGCCAAATATA 2104
                                     :::          :::|||   |
 771 ............................PheArgProAspSerLysGlnT 778

2105 CGGGTTATCTATGCCGCAGCAACACAGTCCTCATCTGTAAAGAACCAACT 2154
     ||                  :::|||||||   :::          :::|||:::
 778 hr............TyrAsnAsnThrTyrValThrValAlaSerProAla 790

2155 GGTTTA..............................AAGTATGAAAAAGC 2174
        |||                                |||||||||::
 791 ThrLeuThrAsnSerIleIleProTrpHisPheProProTyrGluLysSe 807

2175 C.....................................AAAGAGTGGAGGA 2188
     :                                    :::|||:::||||:
 807 rGlyArgLeuAsnValSerAsnThrIleLysAlaIleAsnGluTyrArgL 824

2189 TACCCTGTGTCAACGCCCAG.............................. 2208
     ::   |||   :::   |||
 824 euLeuCysAsnSerArgGlnAlaAspProAlaSerPheLeuGluPheTyr 840

2209 TGGCTTGGCGACATTCTTCTGGGAAACTTTGAGGCACTGAGGCAGATTCA 2258
     :::||||||||   :::          :::   |||        ::
 841 PheLeuGlyAspProMetProHisPheAsnLysIleLeuSerIleAlaAs 857

2259 GTATAGTCGCTACACGGCATTCAGTCTGCAGGATCCA.............. 2295
     :|||:::   |||   :::          |||
```

```
 857 pTyrAsnMetTyrLeuSerArgArgArgCysAspGluAlaAspValLysI 874

2296 ........TTTGCCCCTACCCAGCATTTAGTTTTAAATCTTTTAGAT... 2334
              ::::::   ::::::::::|||   |||   |||||||:::
 874 leHisArgMetSerHisSerAspGlnLeuGlnLeuTyrLeuLeuGluLeu 890

2334 ..................................................... 2334

 891 GlnSerAspGluSerAsnValGluLysTrpLysThrPheTyrArgIleMe 907

2335 .GCTTGGAGAGTTCCCTTA.............................. 2352
         |||   :::||||||
 907 tGlnTrpAspLeuProLeuAsnAsnGluPheProArgIleLeuLeuProS 924

2353 .....................................AAAGTGTCTGCA 2364
                                          |||   |||
 924 erSerLeuAspIleGlyArgProValValAspArgLysLysLysSerIle 940

2365 GAGTTGTTGATGAGT.................ATAAGACTACCTCCCAA 2396
     :::   :::|||:::                    :::   |||||||::
 941 AspGlnValMetAsnHisIleHisArgMetHisSerGlnArgProProSe 957

2397 ACTGAAACAGAATGAA................................ 2412
     ::::   ::::::::
 957 rMetGlyAsnSerSerThrSerSerGluAlaSerSerThrSerProThrA 974

2413 ........GTAGCTAATGTCCAGCCTTCTTCCCAAAGAGCCAGAATTGAA 2454
             :::|||:::   :::|||:::|||:::|||         :::
 974 snAlaAlaThrAlaThrSerSerProAlaSerAsnArgProThrThrSer 990

2455 CACGTACCACCTCCCACT 2472
        :::   |||||||||
 991 ThrAlaGlnProProThr 996
```

seq_name: pir2:A38346

seq_documentation_block:
ultra-high-sulfur keratin 1 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 28-Jun-1991 #sequence_revision 28-Jun-1991 #text_change 24-Sep-1999
C;Accession: A38346
R;Wood, L.; Mills, M.; Hatzenbuhler, N.; Vogeli, G.
J. Biol. Chem. 265, 21375-21380, 1990
A;Title: Serine-rich ultra high sulfur protein gene expression in murine hair and skin c
A;Reference number: A38346; MUID:91065960
A;Accession: A38346
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-230 <WOO>
A;Cross-references: GB:M37759; NID:g200961; PIDN:AAA40106.1; PID:g200962
C;Superfamily: ultra-high-sulfur keratin

alignment_scores:
Quality: 200.00 Length: 218
Ratio: 2.151 Gaps: 21
Percent Similarity: 42.661 Percent Identity: 31.193

alignment_block:
US-09-664-641-10/rev x A38346 ..

Align seg 1/1 to: A38346 from: 1 to: 230

```
1814 GTGCCGCCATGTGCCTGGATTATCCTTTTCCAGGTGGCCAGCAGTTGC.. 1767
     ||||||   |||::                    :::|||||||||
  56 ValProValCysSer.....................CysSerSerCysGl 65

1766 ....TTATCAGACATCTGCTCTGGATAATCCGCAATTGCAAACACACATC 1721
         ::::::   |||:::|||   :::
  65 yGlyCysGlySerSerCysGlyGlyCysGly..................... 75

1720 CCAATAAGAAGCCTTCTTCTGGAATCTCCACTGCTGGA.........TCA 1680
                 |||   |||   |||   :::|||         |||
```

```
  76 ..............SerCysGlySerSerCysGlyGlyCysGlySerSer 87

1679 TGTCCAAAAAGCTGATGCTGC..................TGAGGCGATGG 1648
     |||   |||      ||||||                       ::::::
  88 CysCysLysProValCysCysCysValProValCysSerCysSerSerCy 104

1647 TGGTGGCTGC......TGTTGCTGC.........TGCTGC......TGCT 1619
      |||||||||      |||||||||         ||||||      ||||
 104 sGlyGlyCysLysProCysCysCysGlnSerSerCysCysLysProCysC 121

1618 GTGAAAATGGATGCGGC...GGCTGCTGGGGGTAAGGTTGCTGAGAGATC 1572
     ||:::::::|||||||||   :::|||          :::|||   :::
 121 ysSerSerGlyCysGlySerSerCysCysGlnSerSerCysCysLysPro 137

1571 TGCTGCTGC.....................TGCTGCTGCTGGAGC..... 1548
     |||||||||                     |||||||||   |||
 138 CysCysCysGlnSerSerCysCysLysProCysCysCysGlnSerSerCy 154

1547 .TGCATTATCTGCTGGGGCTGAAGGTGTAAAACCGGGTGC.......... 1509
      |||      |||            |||::::::|||||
 154 sCysLysProCys.............CysSerSerGlyCysGlySerSerC 167

1508 ........................TGCTGCTGC.........TGCTGC 1494
                              |||||||||         ||||||
 167 ysCysGlnSerSerCysCysLysProCysCysCysGlnSerSerCysCys 183

1493 TGGGCCTGCTGCTGC.........TGCTGTAGCATGTGTGTCTCTGGAGT 1453
           |||||||||         ||||||:::   |||
 184 LysProCysCysCysGlnSerSerCysCysLysProCys........... 196

1452 CACTTTCACTTGGCTAAACAGCACTGCATTGGCATTTGTATGT...CCCT 1406
                                 :::   :::      |||   ||||
 197 .....................CysCysGlnSerSerCysCysLysProC 206

1405 GCTGGCTGTGATTCACCTGCTGTTCTAAATTTTTCGTAGGTGCTGAAAGA 1356
     ||              ||||||:::                  :::
 206 ys..............CysCysGln.................SerSer 211

1355 GTCTGTAAGATATGTGCTACATTTGATGGCCGGTTCATCTGCTGAATGTC 1306
        ||||||   |||:::::::       :::||||||||||:::   ||||
 212 CysCysLysProCysCysSerSerGly.CysGlySerSerCys...CysG 227

1305 AG 1304
     ||
 227 ln 227
```

seq_name: pir2:B38346

seq_documentation_block:
ultra-high-sulfur keratin 2 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 24-Sep-1999
C;Accession: A38660; B38346
R;Wood, L.; Mills, M.; Hatzenbuhler, N.; Vogeli, G.
J. Biol. Chem. 266, 4024, 1991
A;Title: Serine-rich ultra high sulfur protein gene expression in murine hair and ski
A;Reference number: A38660; MUID:91154184
A;Accession: A38660
A;Molecule type: DNA
A;Residues: 1-223 <WO2>
A;Cross-references: GB:M37760; NID:g200963; PIDN:AAA40107.1; PID:g200964
A;Note: this is a correction
R;Wood, L.; Mills, M.; Hatzenbuhler, N.; Vogeli, G.
J. Biol. Chem. 265, 21375-21380, 1990
A;Title: Serine-rich ultra high sulfur protein gene expression in murine hair and ski
A;Reference number: A38346; MUID:91065960
A;Accession: B38346
A;Molecule type: DNA
A;Residues: 1-21,'GGCGSGCGGCGSNCGGCGSSCCKPVCC',22-40,'GSS',44-45,'G',47-48,'S',50,'GS
<WOO>
A;Cross-references: GB:M37759; NID:g200961; PIDN:AAA40106.1; PID:g200962
A;Note: the sequence reported in this paper has been corrected. See A38660
C;Superfamily: ultra-high-sulfur keratin

```
alignment_scores:
            Quality:  198.00              Length:    178
              Ratio:  2.676                Gaps:     14
 Percent Similarity:  41.573   Percent Identity: 30.899

alignment_block:
 US-09-664-641-10/rev x B38346   ..

 Align seg 1/1  to: B38346  from: 1  to: 223

    1775 AGCAGTTGCTTATCAGACATCTGCTCTGGATAATCCGCAATTGCAAACAC 1726
         |||||||||   :::      |||         |||:::
      77 SerSerCysCysLysProCysCysCysGlnSerSerCysCys........ 90

    1725 ACATCCCAATAAGAAGCCTTCTTCTGGAATCTCCACTGCTGGATCATGTC 1676
                     ||||||      :::   :::   ::::::||||||
      91 .............LysProCysCysSerSerGlyCysGlySerSerCysC 103

    1675 CAAAAAGCTGATGCTGCTGAGGCGATGGTGGTGCCTGCTGTTGC...... 1632
           :::|||   ||||||                  |||||||||
     103 ysGlnSerSerCysCys.............LysProCysCysCysGlnSer 115

    1631 TGCTGCTGC......TGCTGTGAAAATGGATGCGGC...GGCTGCTGGGG 1591
           ||||||      ||||||:::::|||||||||   :::|||
     116 SerCysCysLysProCysCysSerSerGlyCysGlySerSerCysCysGl 132

    1590 GTAAGGTTGCTGAGAGATCTGCTGCTGC.....................T 1562
           :::|||   :::   |||||||||                      |
     132 nSerSerCysCysLysProCysCysCysGlnSerSerCysCysLysProC 149

    1561 GCTGCTGCTGGAGC......TGCATTATCTGCTGGGGCTGAAGGTGTAAA 1518
         ||||||||   |||      |||      |||           |||:::
     149 ysCysCysGlnSerSerCysCysLysProCys............CysSer 161

    1517 ACCGGGTGC.....................................TGCTG 1504
         :::||||||                                    |||||
     162 SerGlyCysGlySerSerCysCysGlnSerSerCysCysLysProCysCy 178

    1503 CTGC.........TGCTGCTGGGCCTGCTGCTGC.........TGCTGTA 1472
         ||||         ||||||      |||||||||         ||||||:
     178 sCysGlnSerSerCysCysLysProCysCysCysGlnSerSerCysCysL 195

    1471 GCATGTGTGTCTCTGGAGTCACTTTCACTTCGCTAAACAGCACTGCATTG 1422
         ::   |||                                      :::
     195 ysProCys................................CysCysGln 200

    1421 GCATTTGTATGT...CCCTGCTGGCTGTGATTCACCTGCTGTTCTAAATT 1375
         :::      |||   ||||||           :::|||||||:::
     201 SerSerCysCysLysProCysCysCysGlnSerSerCysCysAlaPro.. 216

    1374 TTTCGTAGGTGCTGAAAGAGTCTGTAAGATATGT 1341
                         ||||||      |||
     217 ..................ValCysCysGlnCys 221

seq_name: pir2:T14004

seq_documentation_block:
LrfA protein - slime mold (Dictyostelium discoideum)
C;Species: Dictyostelium discoideum
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000
C;Accession: T14004
R;Saito, J.; Adachi, H.; Sutoh, K.
J. Biol. Chem. 273, 24654-24659, 1998
A;Title: Dictyostelium TRFA homologous to yeast Ssn6 is required for normal growth and e
A;Reference number: Z17852; MUID:98406112
A;Accession: T14004
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1390 <SAI>
A;Cross-references: EMBL:AB009080; NID:d1228566; PID:d1034109; PIDN:BAA33143.1
C;Genetics:
A;Gene: trfA
A;Introns: 333/3; 364/3; 637/1


alignment_scores:
            Quality:  196.50              Length:    265
              Ratio:  1.456                Gaps:     10
 Percent Similarity:  50.943   Percent Identity: 27.170

alignment_block:
 US-09-664-641-10 x T14004   ..

 Align seg 1/1  to: T14004  from: 1  to: 1390

    1207 ATGGCTGCTGGACAAAACCTCCAAAGTTCTGAAAGATCAGAAATGATAGC 1256
         |||||||||   |||:::   :::   ::::::::::::::::::|||||
      12 MetAlaAlaGlnGlnGlnSerMetValAlaGlnGlnGlnGlnGlnIleAl 28

    1257 TACCTGGAGTCCAGCTGTACGGACACTGAGGAATATTACTAATAATGCTG 1306
         |     :::   |||:::
      28 aGlnGlnGlnSerAlaIleAlaGln........................... 36

    1307 ACATTCAGCAGATGAACCGGCCATCAAATGTAGCACATATCTTACAGACT 1356
               ||||||:::   :::   :::::::::|||:::
      37 .....GlnGlnIleAlaGlnGlnGlnGlnIleAlaGln............. 47

    1357 CTTTCAGCACCTACGAAAAATTTAGAACAGCAGGTGAATCACAGCCAGCA 1406
                                  :::|||||||:::   ::::::|||||
      48 .........................GlnGlnGlnIleAlaGlnGlnGlnGl 56

    1407 GGGACATACAAATGCCAATGCAGTGCTGTTTAGCCAAGTGAAAGTGACTC 1456
         |   |||                        :::|||   :::
      56 nGlnHis.......................GlnGlnHisGlnGlnHisG 65

    1457 CAGAGACACACATGCTACAGCAGCAGCAGGCCCAGCAGCAGCAGCAC 1506
           :::   |||:::   |||||||||:::|||   |||:::|||||||||
      65 lnGlnGlnHisGlnGlnGlnGlnGlnHisGlnGlnGlnHisGlnGlnGln 81

    1507 CAGCAC..............CCGGTTTTACACCTTCAGCCCCAGCAGAT 1541
         ||||||                          :::   ||||||||||||
      82 GlnHisIleGlnGlnGlnGlnGlnProGlnGlnGlnGlnProGlnGlnGl 98

    1542 AATGCAGCTCCAGCAGCAGCAGCAGCAGATCTCTCAGCAACCTTACC 1591
         ::::::   ||||||||||||||||:::|||   :::||||||||||
      98 nGlnSerGlnGlnGlnGlnGlnGlnHisGlnGlnGlnGlnProGlnG 115

    1592 CCCAGCAGCCGCCGCATCCATTT........................... 1614
           ||||||||||||:::
     115 lnGlnGlnProProGlnGlnGlnGlnGlnProProGlnGlnGlnGln 131

    1615 ......TCACAGCAGCAGCAGCAGCAACAGCAGCCACCACCATCGCCTCA 1658
               :::|||||||||||||||||||||||||          :::||||||
     132 GlnProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProGl 148

    1659 GCAGCATCAGCTTTTTGGACATGATCCAGCAGTGGAGATTCCAGAAGAAG 1708
         ||||:::|||                          :::      ::::::
     148 nGlnGlnGln.....................GlnGlnGlnGlnGlnA 157

    1709 GCTTCTTATTGGGATGTGTGTTTGCAATTGCCGAT............... 1743
               |||   |||   :::      |||   |||
     157 spGlnLeuSerGlyValThrGlyLysIleArgAspLeuAsnGluSerIle 173

    1744 .................TATCCAGAGCAGATGTCTGATAAG...CAACT 1772
                         |||   |||:::::::::::::::   :::
     174 TrpIleHisLeuGlyGlyTyrAlaGluSerIleGlyGluGlnAspLysAl 190

    1773 GCTGGCCACCTGGAAAAGGATAATCCAGGCACATGGCGGCACTGTTGACC 1822
         ||||||:::::::::::::   ::::::   |||              :::|
     190 aLeuAlaSerTyrGluAsnAlaLeuArg...His............AsnP 202

    1823 CCACCTTCACGAGTCGATGCACGCACCTTCTCTGTGAGAGTCAAGTCAGC 1872
         ||          :::      |||::::::           :::::::::
     202 roPheSerIleLysAlaLeuThrGlnIleAlaSerLeuPheArgIleLys 218
```

```
    1873 AGCCCGTATGCACAGGCAATAACAGAAAGAAAGACATGTCTTACT 1917
          :::   |||::::::|||   :::      :::|||   ||||||
     219 GluGlnTyrSerLysAlaAlaGluTyrPheGlnArgIleValThr 233

seq_name: pir2:T03455

seq_documentation_block:
ALR protein - human
C;Species: Homo sapiens (man)
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 21-Jul-2000
C;Accession: T03455
R;Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Yano,
Oncogene 15, 549-560, 1997
A;Title: Structure and expression pattern of human ALR, a novel gene with strong homolog
A;Reference number: Z14954; MUID:97388474
A;Accession: T03455
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-4957 <PRA>
A;Cross-references: EMBL:AF010404; NID:g2358286; PIDN:AAC51735.1; PID:g2358287
C;Genetics:
A;Gene: ALR
A;Map position: 12
C;Superfamily: human ALR protein
C;Keywords: alternative splicing


alignment_scores:
            Quality:  196.50            Length:    375
              Ratio:  1.062               Gaps:     15
 Percent Similarity:  49.333   Percent Identity: 25.067

alignment_block:
 US-09-664-641-10 x T03455  ..

 Align seg 1/1  to: T03455  from: 1  to: 4957

     835 GAAGAGGAAGAAGAGGAAGAGGAAGAGGAGGAG.....GAAGT 872
         :::::::::::::::::::::::::::::::::           :::||
    3155 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnHisLeuLeuGlyGlnVa 3171

     873 AGAAAATGAGGAACAAGATTCTCAGAATGAGGGT...AGTACAGATGAGA 919
         |      ::::::|||:::::::|||:::   |||   :::|||::::::
    3171 lAlaIleGlnGlnGlnGlnGlnGlnGlyProGlyValGlnThrAsnGlnA 3188

     920 AGTCAAGCCCTGCCAGCTCTCAAGAAGGGTCTCCTTCAGGTGAC..... 963
           :::|||      :::          ||||||:::
    3188 laLeuGlyProLysProGlnGlyLeuMetProProSerSerHisGlnGly 3204

     964 .........CAGCAGTTTTCACCTAAATCCAACACTGAAAAATCTAAAGG 1004
                 ||||||:::|||||||:::
    3205 LeuLeuValGlnGlnLeuSerProGln..................... 3213

    1005 GGAATTAATGTTTGATGATTCTTCAGATTCATCACCGGAAAAACAGGAGA 1054

    3213 ..................................................... 3213

    1055 GAAATTTAAACTGGACCCCGGCCGAAGTCCCACAG......TTAGCTGCA 1098
                       |||   :::   ||||||      |||:::
    3214 .................ProProGlnGlyProGlnGlyMetLeuGlyPro 3224

    1099 GCAAAACGCAGGCTGCCTCAGGGAAAGGAGCCTGGGTTGATTAACTTGTG 1148
         |||:::      :::   |||   ::::::||||||   ::::::
    3225 AlaGlnValAlaValLeuGlnGlnGlnHisProGlyAlaLeuGly..... 3239

    1149 TGCCAATGTCCCACCCGTCCCAGGTAACATTTTGCCCCCTGAGGTCCGGG 1198
                  |||      |||                     |||
    3240 ..........ProGlnGlyProHis.....................ArgG 3246

    1199 GTAATTTAATGGCTGCTGGACAAAACCTCCAAAGTTCTGAAAGATCAGAA 1248
           ||||||:::   ::::::   |||:::|||   :::
    3246 lnValLeuMetThrGlnSerArgValLeuSerSerProGln......... 3259

    1249 ATGATAGCTACCTGGAGTCCAGCTGTACGGACACTGAGGAATATTACTAA 1298
            :::|||      :::   ::::::          |||   :::|||
    3260 ...LeuAlaGlnGlnGlyGlnGlyLeuMetGlyHisArgLeuValThrAl 3275

    1299 TAATGCTGACATTCAGCAGATGAACCGGCCATCAAATGTAGCACATATCT 1348
          :::   :::   ||||||::::::                    |||
    3275 aGlnGlnGlnGlnGlnGlnGlnGln.................HisGlnG 3286

    1349 TACAGACTCTTTCAGCACCTACGAAAAATTTAGAACAGCAGGTGAATCAC 1398
           |||         |||      ::::::|||:::|||::::::
    3286 lnGlnGlySerMetAlaGlyLeuSerHisLeuGlnGlnSerLeu...Met 3301

    1399 AGCCAGCAGGGACATACAAATGCCAATGCAGTGCTGTTTAGCCAAGTGAA 1448
         |||:::::::|||:::   :::   :::|||      ::::::::::::::
    3302 SerHisSerGlyGlnProLysLeuSerAlaGlnProMetGlySerLeuGl 3318

    1449 AGTGACTCCAGAGACACACATGCTACAGCAGCAGCAGCAGGCCCAGCAGC 1498
         :       :::   ::::::||||||||||||||||||||   |||||||
    3318 nGlnLeuGlnGlnGlnGlnGlnLeuGlnGlnGlnGlnGlnLeuGlnGlnG 3335

    1499 AGCAGCAGCAG......CACCCGGTTTTACACCTTCAGCCCCAGCAGATA 1542
         ||||||||||      :::         :::||||||   ||||||:::
    3335 lnGlnGlnGlnGlnLeuGlnGlnGlnGlnGlnLeuGlnGlnGlnGlnLeu  351

    1543 ATGCAGCTCCAGCAGCAGCAGCAGCAGCAGATCTCTCAGCAACCTTACCC 1592
         :::|||   ||||||||||||||||   |||   :::||||||
    3352 GlnGlnGlnGlnGlnGlnGlnGlnLeuGlnGlnGlnGlnGlnGlnGlnLe 3368

    1593 CCAGCAGCCGCCGCATCCATTTTCACAGCAGCAGCAGCAGCAACAGCAGC 1642
          ||||||      :::   ::::::||||||||||||||||||||||||
    3368 uGlnGlnGlnGlnGlnGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnG 3385

    1643 CACCACCATCGCCTCAGCAGCATCAGCTTTTTGGACATGATCCAGCAGTG 1692
               :::   ||||||:::|||   :::|||
    3385 lnPheGlnGlnGlnGlnGlnGlnGlnMetGly..................... 3396

    1693 GAGATTCCAGAAGAAGGCTTCTTATTGGGATGTGTGTTTGCAATTGCGGA 1742
            :::   ::::::      |||||||:::
    3397 LeuLeuAsnGlnSerArgThrLeuLeuSer..................... 3406

    1743 TTATCCAGAGCAGATGTCTGATAAGCAA..................... 1770
             |||:::|||:::::::::::::|||
    3407 ....ProGlnGlnGlnGlnGlnGlnGlnValAlaLeuGlyProGlyMetP 3422

    1770 ..................................................... 1770

    3422 roAlaLysProLeuGlnHisPheSerSerProGlyAlaLeuGlyProThr 3438

    1771 CTGCTGGCCACCTGGAAAAGGATAATCCAGGCACATGGCGGCACTGTTGA 1820
         ||||||   |||   |||                  :::   :::|||||||||
    3439 LeuLeuLeuThrGlyLys..............GluGlnAsnThrValAs 3450

    1821 CCCCACCTTCACGAGTCGATGCACG 1845
         ||||:::   :::|||::::::|||
    3450 pProAlaValSerSerGluAlaThr 3458
```

OM of: US-09-664-641-10 to: Issued_Patents_AA:* out_format : pfs

Date: Jun 4, 2002 5:33 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09664641/runat_04062002_110932_9169/app_query.fasta_1.3697
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=rai -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09664641_@CGN1_1_44 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-664-641-10
Query length: 3580
Database: Issued_Patents_AA:*
Database sequences: 231628
Database length: 24425594
Search time (sec): 55.080000

score_list:

Sequence	Strd	Orig	%Score	EScore	Len	! Documentation ..
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-977-767-3	-		236.50	356.91	1.6e-13	1345 !
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-630-915A-37	-		236.00	355.54	1.8e-13	1400 !
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-900-230-3	-		224.50	335.72	2.3e-12	1417 !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-728-323A-2	+		214.00	320.29	2.0e-11	1162 !
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-415-751-6	-		197.00	306.02	4.0e-10	362 !
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-07-906-349A-6	-		196.00	294.23	8.2e-10	801 !
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-415-751-35	-		194.00	300.89	7.7e-10	362 !
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-07-814-964-13	+		184.00	278.67	9.0e-09	542 !
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-258-442-13	+		184.00	278.67	9.0e-09	542 !
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-328-809-8	+		184.00	278.67	9.0e-09	542 !
/cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:PCT-US92-11107-13	+		184.00	278.67	9.0e-09	542
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-990-823-112	-		182.50	293.02	5.4e-09	143 !
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-916-352-2	+		174.00	253.74	1.2e-07	1004 !
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-508-761B-22	-		173.00	279.00	3.9e-08	120 !
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-625-188-20	+		173.00	256.09	1.2e-07	729 !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-853-310-2	+		173.00	253.37	1.4e-07	903 !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-267-803B-9	+		169.00	247.82	3.1e-07	816 !
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-041-886-17	+		169.00	247.82	3.1e-07	816 !
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-910-925-3	+		162.50	237.89	1.2e-06	743 !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-991-300-2	+		159.50	241.65	1.5e-06	369 !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-918-914-4	+		158.00	229.45	3.4e-06	788 !
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-115-446-2	+		153.00	221.85	9.7e-06	731 !
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-041-886-28	+		152.50	225.49	8.7e-06	513 !
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-041-886-29	+		152.50	225.08	8.8e-06	530 !
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-041-886-30	+		152.50	224.56	9.1e-06	552 !
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-041-886-31	+		152.50	223.74	9.4e-06	589 !
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-961-083-56	+		152.50	219.92	1.1e-05	796 !
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-246-982A-6	+		152.50	202.47	2.7e-05	3144 !
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-453-265-6	+		152.50	202.47	2.7e-05	3144 !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-457-273B-42	+		152.50	202.47	2.7e-05	3144 !
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-556-419-21	+		152.50	202.47	2.7e-05	3144 !
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-041-886-15	+		152.50	202.47	2.7e-05	3144 !
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-227-536-2	+		151.50	204.12	2.9e-05	2414 !
/cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:PCT-US95-04682-2	+		151.50	204.12	2.9e-05	2414
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-910-925-4	+		151.00	218.93	1.5e-05	703 !
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-194-468-2	+		150.50	202.26	3.6e-05	2441 !
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-961-739-2	+		150.50	202.26	3.6e-05	2441 !
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-185-432-2	+		149.50	215.76	2.1e-05	737 !
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-185-432-4	+		149.50	215.76	2.1e-05	737 !
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-910-925-1	+		149.00	215.26	2.3e-05	717 !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-785-310A-7	+		149.00	213.49	2.5e-05	824
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-816-693A-2	+		148.50	212.17	2.9e-05	855
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-885-291-2	+		148.50	212.17	2.9e-05	855
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-496-672-2	+		148.50	212.17	2.9e-05	855
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-625-188-10	+		147.50	214.02	3.0e-05	646

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-977-767-3

seq_documentation_block:
; Sequence 3, Application US/08977767
; Patent No. 5972684
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Yue, Henry
; APPLICANT: Greenwald, Sara
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: CARBONIC ANHYDRASE VIII
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/977,767
; FILING DATE: Herewith
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0423 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1345 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1532042
US-08-977-767-3

alignment_scores:
Quality: 236.50 Length: 194
Ratio: 2.489 Gaps: 13
Percent Similarity: 48.969 Percent Identity: 32.990

alignment_block:
US-09-664-641-10/rev x US-08-977-767-3 ..

Align seq 1/1 to: US-08-977-767-3 from: 1 to: 1345

1778 GCCAGCAGTTGCTTATCAGACATCTGCTCTGGATAATCCGCAATTGCAAA 1729
1057 AlaAlaThrCysCysThrCysCysGlyGlyGlyThrCysCysCysAlaGl 1073

1728 CACACATCCCAATAAGAAGCCTTCTTCTGGAATCTCCACTGCTGGATCAT 1679

```
     1073 yThrCysThrThrCysCysAlaGlyAlaGlyCysThrCysAlaGlyCysC 1090

     1678 GTCCAAAAAGCTGATGCTGCTGAGGCGATGGTGGTGGCTGCTGTTGCTGC 1629
          ||      :::   :::|||   |||      ||||||||||||||||||
     1090 ysGlyThrAla...AlaCysGlyGlyCysCysGlyGlyCysCysCysCys 1105

     1628 ......TGC......TGCTGCTGTGAAAATGGATGCGGCGGCTGC..... 1596
                |||      |||||||||   :::   ||||||||||||
     1106 ThrGlyCysAlaGlyCysCysCysThrThrCysCysGlyGlyCysAlaCy 1122

     1595 ..............................TGGGGGTAAGC TGCTGAG 1577
                                          |||   ||||||
     1122 sAlaGlyGlyGlyCysCysThrThrGlyAlaGlyGlyGlyGlyCys...A 1138

     1576 AGATCTGC............TGCTGCTGCTGCTGCTGGAGCTGCATTATC 1539
               |||            ||||||||||||||   :::|||
     1138 laAlaCysAlaGlyGlyAlaCysCysCysCysCysGlyGlyCysCysAla 1154

     1538 TGC.......................TGGGGCTGAAGGTGT........ 1521
          |||                        |||      |||
     1155 CysCysCysGlyAlaGlyAlaGlyGlyGlyGlyThrGlyCysGlyGlyAl 1171

     1520 .AAAACCGGGTGCTGCTGCTGCTGCTGCTGGGCCTGCTGCTGCTGCTGTA 1472
             ||||||||||||:::::::|||   |||||||||   ||||||:
     1171 aGlyGlyGlyCysCysCysAlaAlaCysThrAlaCysCysGlyCysCysT 1188

     1471 GCATGTGTGTCTCTGGAGTCACTTTCACTTGGCTAAACAGCACTGCATTG 1422
          ::   |||:::::::|||:::               ::::::|||:::
     1188 hrGlyCysAlaThrGlyThrGly............GlyAlaThrGlyAla 1200

     1421 GCATTTGTATGTCCCTGCTGGCTGTGATTCACCTGCTGTTCTAAA..... 1377
          :::   :::|||   |||                |||||||:::
     1201 ThrGlyThrCysCysCys..............CysCysAlaThrGlyGl 1212

     1376 .......TTTTTCGTAGGTGCTGAAAGA...GTCTGTAAGATATGTGCTA 1337
                  :::||||||        :::|||      |||||||
     1212 yThrCysThrCysThrGlyAlaGlyAlaCysThrCysCysCysCysAlaT 1229

     1336 CATTTGATGGCCGGTTCATCTGCTGAATGT 1307
          ||       ||||||:::::::::   |||
     1229 hrCysGly.AlaGlyGlyAlaGlyThrCys 1238

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-630-915A-37

seq_documentation_block:
; Sequence 37, Application US/08630915A
; Patent No. 6309820
;  GENERAL INFORMATION:
;    APPLICANT:  SPARKS, Andrew B.
;    APPLICANT:  HOFFMAN, No. 6309820h
;    APPLICANT:  KAY, Brian K.
;    APPLICANT:  FOWLKES, Dana M.
;    APPLICANT:  McCONNELL, Stephen J.
;    TITLE OF INVENTION:  POLYPEPTIDES HAVING A FUNCTIONAL
;    TITLE OF INVENTION:  DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
;    TITLE OF INVENTION:  USING SAME
;    NUMBER OF SEQUENCES:  227
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/630,915A
;      FILING DATE:  03-APR-1996
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  1101-174
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  37:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1400 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-08-630-915A-37


alignment_scores:
            Quality:  236.00              Length:    228
              Ratio:  2.408                 Gaps:     14
 Percent Similarity:  42.982   Percent Identity: 28.947

alignment_block:
 US-09-664-641-10/rev x US-08-630-915A-37  ..

 Align seg 1/1  to: US-08-630-915A-37  from: 1  to: 1400

    1935 TGTGTTTAACCAGTGTGCAGTAACACATCTCTTTCTTTCTCTTATTGCCT 1886
         |||:::          |||:::::::::::   :::   :::   |||
     337 CysThrAlaThrGlyCysAlaGlyAlaGlyAlaAlaAlaAlaThrCysCy 353

    1885 GTGCATACGCGCTGCTGACTTGACTCTCACAGAGAAGGTGCGTGCATCGA 1836
          |||      ||||||                              |||:::|
     353 sCysAlaGlyAlaAla.........................AlaAlaT 361

    1835 CTCGTGAAGGTGGGGTCAACAGTGCCGCCATGTGCCTGGATTATCCTTTT 1786
         ||      ||||||::::::   ::::::   |||
     361 hrGlyAlaGlyGlyThrThrCysCysCysGlyCysThrCysCysAlaGly 377

    1785 CCAGGTGGCCAGCAGTTGCTTATCAGACATCTGCTCTGGATAATCCGCAA 1736
         ||| |||:::::|||           ||||:::|||   :::|||
     378 ThrGly.AlaAlaAlaCys...........CysAlaGlyThrGlyAla. 389

    1735 TTGCAAACACACATCCCAATAAGAAGCCTTCTTCTGGAATCTCCACTGCT 1686
                                        :::|||   :::||||:::
     390 ..........................CysThrGlyAlaThrThrCys 396

    1685 GGATCATGTCCAAAAAGCTGATGCTGCTGAGGCGATGGTGGTGGCTGCTG 1636
         ::::::|||              |||                |||||||
     397 AlaAlaCysAlaThr......CysThr.............GlyCysCy 406

    1635 TTGC.........TGCTGCTGCTGCTGTGAAAATGGATGC...GGCGGCT 1598
         ||||         ||||||||||||||||      :::|||   |||||||
     406 sCysCysThrGlyCysCysCysCysCysAlaAlaAlaCysThrGlyGlyC 423

    1597 GCTGGGGGTAAGGTTGCTGAGAG.........ATCTGCTGCTGCTGCTGC 1557
         ||         ||||||                |||||||||||||||
     423 ysCysThrThrGlyCysGlyThrGlyAlaGlyAlaCysCysCysCysCys 439

    1556 ..............................................TG 1555
                                                       ||
     440 GlyCysCysCysCysThrThrThrGlyGlyCysAlaGlyThrAlaAlaCy 456

    1554 CTGGAGCTGCATTATCTGCTGGGGCTGAAGGTGTAAAACCGGGTGCTGC. 1506
         |   :::|||      |||   |||            :::|||||||||
     456 sCysThrCysThrThrCysAlaGly............AlaGlyCysCysC 469

    1505 .....TGCTGCTGCTGCTGGGCCTGCTGCTGCTGC.........TGTAGC 1470
          ||||||:::|||   ||||||||||||||||                :::
     469 ysThrCysCysAlaCysGlyAlaCysCysCysCysThrAlaAlaThrAla 485

    1469 ATGTGTGTCTCTGGAGTCACTTTCACTTGGCTAAACAGCACTGCATTGGC 1420
```

```
          |||::::::|||                  ::::::  |||  ::
     486 AlaCysThrGlyGlyGlyCysCysGlyAlaCysThrThrCysAlaGlyCy 502

    1419 ATTTGTATGTCCCTGCTGGCTGTGATTCACCTGCTGTTCTAAATTTTTCG 1370
         :     |||  |||              |||||||:::
     502 sThrCysCysAlaCysGlyThrGlyGlyCysCysCysAla..... 515

    1369 TAGGTGCTGAAAGAGTCTGTAAGATATGTGCT 1338
                        |||     ||||||
     516 .............CysCysAlaGlyCysAla 521

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-900-230-3

seq_documentation_block:
; Sequence 3, Application US/08900230
; Patent No. 6329197
;  GENERAL INFORMATION:
;    APPLICANT:  Bard, Jonathan A.
;    TITLE OF INVENTION:  DNA ENCODING GALANN GALR3 RECEPTORS AND
;    TITLE OF INVENTION:  USES THEREOF
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLp
;      STREET:  1185 Avenue of The Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  11036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/900,230
;      FILING DATE:  23-JUL-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  52241-C/JPW/ADM
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-278-0400
;      TELEFAX:  212-391-0525
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1417 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-900-230-3


alignment_scores:
            Quality:  224.50              Length:    150
              Ratio:  3.454                 Gaps:     10
 Percent Similarity:  43.333   Percent Identity: 36.667

alignment_block:
 US-09-664-641-10/rev x US-08-900-230-3   ..

 Align seq 1/1  to: US-08-900-230-3  from: 1  to: 1417

    1685 GGATCATGTCCAAAAAGCTGATGCTGCTGAGGCGATGGTGGTGGCTGC.. 1638
         |||  |||      :::  ||||||  |||  |||  ||||||
    1064 GlyCysCysAlaCysThrThrCysCys...GlyCysGlyCysGlyCysGl 1079

    1637 .................TGTTGCTGCTGCTGC.............T 1622
                          ||||||  ||||||              |
    1079 yCysThrThrCysCysGlyCysCysGlyCysCysThrGlyThrGlyGlyC 1096

    1621 GCTGTGAAAATGGATGCGGCGGCTGCTGGGGG.............TAAGGT 1584
         |||||  :::|||||||||||||||  |||              |||
    1096 ysCysGlyThrGlyCysGlyGlyCysCysGlyCysCysGlyAlaCysGly 1112

    1583 TGCTGAGAGATCTGCTGCTGCTGC.........TGCTGCTGGAGCTGCAT 1543
         |||         |||:::||||||        ||||||  :::|||
    1113 CysCysGlyCysCysAlaCysCysGlyThrGlyCysCysCysGlyCysCy 1129

    1542 TATCTGCTGGGGCTGAAGGTGTAAAACCGGGTGC.............. 1509
           |||           |||  |||||||||
    1129 sGlyCysGly......CysCysThrThrGlyCysGlyThrCysGlyCysG 1144

    1508 .............TGCTGCTGCTGC...TGCTGGGCCTGCTGC..... 1482
                      ||||||||||||  |||  :::||||||
    1144 lyThrCysCysGlyCysCysCysCysGlyCysGlyThrCysCysThrCys 1160

    1481 .........TGCTGCTGTAGCATGTGTGTCTCTGGAGTCACTTTCACTTG 1441
                  ||||||||||:::  |||  :::|||  |||
    1161 GlyGlyGlyCysCysCysAlaCysCysCysGlyGlyCysThrGlyCysCy 1177

    1440 GCTAAACAGCACTGCATTGGCATTTGTATGTCCCTGCTGGCTGTGATTCA 1391
              :::  |||  |||      |||  |||
    1177 sCysCysGlyGlyAlaGlyAlaCysGlyCysCysCysGly.........G 1191

    1390 CCTGCTGTTCTAAATTTTTCGTAGGTGCTGAAAGAGTCTGTAAGATATGT 1341
           ||||||:::           |||:::      |||      |||
    1191 lyCysCysThrAlaGlyCysGlyGlyGlyAlaGlyGlyCysThrGlyCys 1207

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-728-323A-2

seq_documentation_block:
; Sequence 2, Application US/08728323A
; Patent No. 5948676
;  GENERAL INFORMATION:
;    APPLICANT:  Chang, Yuan
;    APPLICANT:  Bohenzky, Roy A.
;    APPLICANT:  Russo, James J.
;    APPLICANT:  Edelman, Isidore S.
;    APPLICANT:  Moore, Patrick S.
;    TITLE OF INVENTION:  Immediate Early Protein From Kaposi's
;    TITLE OF INVENTION:  Sarcoma-Associated Herpesvirus, DNA
;    TITLE OF INVENTION:  Encoding Same And Uses Thereof
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/728,323A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  0575/52268/JPW/MSC/SKS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-278-0400
;      TELEFAX:  212-391-0525
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH: 1162 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
```

```
US-08-728-323A-2


alignment_scores:
            Quality:  214.00            Length:    317
              Ratio:  1.417               Gaps:     11
 Percent Similarity:  47.634   Percent Identity: 24.606

alignment_block:
 US-09-664-641-10 x US-08-728-323A-2  ..

 Align seg 1/1  to: US-08-728-323A-2  from: 1  to: 1162

     835 GAAGAGGAAGAAGAGGAAGAGGAAGACGAGGACGAACTAGAAAATGAGGA 884
         ||||||||||:::||||||||:::|||||||||||  |||:::::::::
     390 GluGluGluAspGluGluGluAspGluGluGluAspGluGluAspAs 406

     885 ACAAGATTCTCAGAATGAGGGT...ACTACAGATGAGAAGTCAAGCCCTG 931
         ::::|||::::::|||||      :::  ||||||:::::::::
     406 pAspAspGluAspAsnGluAspGluGluAspAspGluGluGluAspLysL 423

     932 CCAGCTCTCAAGAAGGGTCTCCTTCAGGTGACCAGCAGTTTTCACCTAAA 981
           :::::::::|||  :::  :::|||:::::  :::|||  :::
     423 ysGluAspGluGluAspGlyGlyAspGlyAsnLysThrLeuSerIleGln 439

     982 TCCAACACTGAAAAATCTAAAGGGGAATTAATGTTTGATGATTCTTCAGA 1031
         |||:::  :::::::::     |||  :::  ::::::  :::::
     440 SerSerGlnGlnGlnGln......GluProGlnGlnGlnGluProGlnGl 454

    1032 TTCATCACCGGAAAAACAGGAGAGAAATTTAAACTGGACCCCGGCCGAAG 1081
         ::::::::|||:::::||||||                |||  :::
     454 nGlnGluProGlnGlnGlnGlu.................ProLeuGlnG 465

    1082 TCCCACAGTTAGCTGCAGCAAAACGCAGGCTGCCTCAGGGAAAGGAGCCT 1131
           ||||||           :::::::::  ||||||  :::||||||
     465 luProGlnGlnGlnGluProGlnGlnGlnGluProGlnGlnGlnGluPro 481

    1132 GGGTTGATTAACTTGTGTGCCAATGTCCCACCCGTCCCAGGTAACATTTT 1181

     481 ..................................................  481

    1182 GCCCCCTGAGGTCCGGGGTAATTTAATGGCTGCTGGACAAAACCTCCAAA 1231
                                                  ||||||:
     482 .....................................LeuGlnG 484

    1232 GTTCTGAAAGATCAGAAATGATAGCTACCTGGAGTCCAGCTGTACGGACA 1281
         ::   ::::::::::|||
     484 luProGlnGlnGlnGlu.............................. 489

    1282 CTGAGGAATATTACTAATAATGCTGACATTCAGCAGATGAACCGGCCATC 1331
                                    |||:::::::::|||
     490 ............................ProGlnGlnGlnGluPro.. 495

    1332 AAATGTAGCACATATCTTACAGACTCTTTCAGCACCTACGAAAAATTTAG 1381
                               |||:::  |||  ::::::
     496 .......................LeuGlnGluProGlnGlnGlnGluP 504

    1382 AACAGCAGGTGAATCACAGCCAGCAGGGACATACAAATGCCAATGCAGTG 1431
           ||||||  :::  :::|||||||
     504 roGlnGlnGlnGluProGlnGlnGln..................... 512

    1432 CTGTTTAGCCAAGTGAAAGTGACTCCAGAGACACACATGCTACAGCAGCA 1481
                        .    |||:::  :::      ||||||||
     513 ....................GluProGlnGlnGlnGluProGlnGlnGl 522

    1482 GCAGCAGGCCCAGCAGCAGCAGCAGCACCCCGGTTTTACACCTTCAGC 1531
         |:::     ||||||:::  ||||||::                :::|
     522 nGluProGlnGlnGlnGluProGlnGlnGln.............GluP 534

    1532 CCCAGCAGATAATGCAGCTCCAGCAGCAGCAGCAGCAGATCTCTCAG 1581
         ||||||||           ||||||:::  ||||||||     |||
     534 roGlnGlnGlnGluProGlnGlnGlnGluProGlnGlnGlnGluProGln 550

    1582 CAACCTTACCCCCAGCAG......CCGCCGCATCCATTTTCACAGCAGCA 1625
         |||      |||||||||      |||  :::         ||||||::
     551 GlnArgGluProGlnGlnArgGluProGlnGlnArgGluProGlnGlnAr 567

    1626 GCAGCAGCAACAGCAGCCACCACCA......TCGCCTCAGCAGCATCAGC 1669
         ::::  ||||||:::  |||         :::||||||||||::::::
     567 gGluProGlnGlnArgGluProGlnGlnArgGluProGlnGlnArgGluP 584

    1670 TTTTTGGACATGATCCAGCAGTGGAGATTCCAGAAGAAGGCTTCTTATTG 1719
               ::::::|||      :::  |||::::::
     584 roGlnGlnArgGluProGlnGlnArgGluProGlnGln............ 596

    1720 GGATGTGTGTTTGCAATTGCGGATTATCCAGAGCAGATGTCTGATAAGCA 1769
                                          ||||||::::::|||:::||
     597 .......................GlnAspGluGlnGlnGlnAspGluGl 605

    1770 A 1770
         |
     605 n 605

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-415-751-6

seq_documentation_block:
; Sequence 6, Application US/08415751
; Patent No. 5643772
;  GENERAL INFORMATION:
;    APPLICANT:  PETERSEN, CAROLYN
;    APPLICANT:  LEECH, JAMES
;    APPLICANT:  NELSON, RICHARD, C.
;    APPLICANT:  GUT, JIRI
;    TITLE OF INVENTION:  POLYPEPTIDES BINDING ANTI-
;    TITLE OF INVENTION:  CRYPTOSPORIDIUM ANTIBODIES, DNA
;    TITLE OF INVENTION:  AND RNA ENCODING THEM, HYBRID
;    TITLE OF INVENTION:  VECTOR AND TRANSFORMED HOST AND
;    TITLE OF INVENTION:  METHODS FOR IMMUNOTHERAPY AND
;    TITLE OF INVENTION:  DIAGNOSIS AND KIT
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PHILLIPS, MOORE, LEMPIO & FINLEY
;      STREET:  385 Sherman Avenue, Suite 6
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94306-1840
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette - 3.5 inch, 1.44 Kb storage
;      COMPUTER:  PC
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  Wordperfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/415,751
;      FILING DATE:  03-APR-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/071,880
;      FILING DATE:  June 1, 1993
;      APPLICATION NUMBER:  07/891,301
;      FILING DATE:  May 29, 1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hana Dolezalova
;      REGISTRATION NUMBER:  30,518
;      REFERENCE/DOCKET NUMBER:  480.19-2 (HHD)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 324-1677
;      TELEFAX:  (415) 324-1678
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  362 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    ORIGINAL SOURCE:
```

```
;      ORGANISM:  Cryptosporidium parvum
;    FEATURE:
;      NAME/KEY:  Positions coded by nonsense codons are
;      NAME/KEY:  identified as Xaa.
US-08-415-751-6


alignment_scores:
            Quality:  197.00            Length:    217
              Ratio:  2.189              Gaps:     16
 Percent Similarity:  41.475   Percent Identity: 26.728

alignment_block:
 US-09-664-641-10/rev x US-08-415-751-6   ..

 Align seq 1/1  to: US-08-415-751-6  from: 1  to: 362

    1896 TCTTATTGCCTGTGCATACGCGCTGCTGACTTGACTCTCACAGAGAAGGT 1847
         ||||||      |||:::|||:::   :::      :::      :::||
     123 SerTyrLeuSerCysLeuArgThrTrpAsnTrpPheIleIle...ArgVa 138

    1846 GCGTGCATCGACTCGTGAAGGTGGGGTCAACAGTGCCGCCATGTGCCTGG 1797
         |   ::::::|||           :::|||   ::::::
     138 l*********ThrThrValIleValIleAsnHisSerCys..... 151

    1796 ATTATCCTTTTCCAGGTGGCCAGCAGTTGCTTATCAGACATCTGCTC... 1750
                               :::||||||::||||||:::
     152 .........................ValLeuIleGlnHisLeuPheAsn 159

    1749 ......TGGATA.............................ATCCGCAA 1736
               |||                                  |||
     160 TrpAspTrpCysAsnPheAsnValThrAspIleArg***TrpCysArgCy 176

    1735 TTGCAAACACACATCCCAATAAGAAGCCTTCTTCTGGAATCTCCACTGCT 1686
          |||   |||                  |||   :::|||    :::
     176 sCysCysHis.....................PheValSerLeu.GlySer 185

    1685 GGATCATGTCCAAAAAGCTGATGCTGCTGAGGCGATGGTGGTGGCTGCTG 1636
         :::||||||          ||||||                  |||||
     186 SerSerCysArgTrpHisCysCysCys.................CysCy 196

    1635 TTGCTGCTGC...TGCTGCTGTGAAAATGGATGCGGCGGCTGCTGGGGGT 1589
         ||||||||||   |||||||||         |||   :::   |||
     196 sCysCysCysPheCysCysCys.........CysCysAsnTyrTrp.... 208

    1588 AAGGTTGCTGAGAGATCTGCTGCTGCTGCTGCTGCTGGAGCTGCATTATC 1539
                  :::   ||||||||||||||||           :::
     209 ..............Leu***CysCysCysCysCysGlyTyrTrpTrpLeu 220

    1538 TGCTGGGGCTGAAGGTGTAAAACCGGGTGCTGCTGCTGCTGCTGCTGGGC 1489
           |||              :::   ||||||||||||||||||||
     221 ThrTrp..............AsnLeuCysCysCysCysCysTrpPh 232

    1488 CTGCTGCTGCTGCTGTAGCATGTGTGTCTCTGGAGTCACTTTCACTTGGC 1439
               |||||||||:::                          |||:
     232 eLeuSerCysCysCysAsn.......................AspTrpI 241

    1438 TAAACAGCACTGCATTGGCATTTGTATGTCCCTGCTGGCTGTGATTCACC 1389
         :::::|||                 |||   |||
     241 leLysSer..............CysCysGlyCys.............. 247

    1388 TGCTGTTCTAAATTTTTCGTAGGTGCTGAAAGAGTCTGTAAGATATGTGC 1339
         ||||||   :::   :::      :::              :::|||
     248 CysCysLeuArg***Trp*********.............LeuValCysLe 260

    1338 TACATTTGATGGCCGGTTCATCTGCTGAATGTCAGCATTATTAGTAATA 1290
           |||:::                 :::::::::||||||||:::
     260 uAspPheGln.................Leu******LeuLeuValVal 270

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-07-906-349A-6

seq_documentation_block:
; Sequence 6, Application US/07906349A

; Patent No. 5434064
; GENERAL INFORMATION:
;   APPLICANT:  Schlessinger, Joseph
;   APPLICANT:  Skolnik, Edward Y.
;   APPLICANT:  Margolis, Benjamin L.
;   TITLE OF INVENTION:  A NOVEL EXPRESSION-CLONING METHOD FOR
;   TITLE OF INVENTION:  IDENTIFYING TARGET PROTEINS FOR EUKARYOTIC TYROSINE KINASE
;   TITLE OF INVENTION:  TARGET PROTEINS
;   NUMBER OF SEQUENCES:  16
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Browdy and Neimark
;     STREET:  419 Seventh Street, N.W.
;     CITY:  Washington
;     STATE:  D.C.
;     COUNTRY:  USA
;     ZIP:  20004
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/07/906,349A
;     FILING DATE:  30-JUN-1992
;     CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  07/643,237
;     FILING DATE:  18-JAN-1991
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  202-628-5197
;     TELEFAX:  202-737-3528
; INFORMATION FOR SEQ ID NO:  6:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  801 amino acids
;     TYPE:  amino acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  protein
US-07-906-349A-6


alignment_scores:
            Quality:  196.00            Length:    420
              Ratio:  1.273              Gaps:     22
 Percent Similarity:  36.667   Percent Identity: 22.381

alignment_block:
 US-09-664-641-10/rev x US-07-906-349A-6   ..

 Align seq 1/1  to: US-07-906-349A-6  from: 1  to: 801

    2257 GAATCTGCCTCAGTGCCTCAAAGTTTCCCA..................... 2228
         ||||||   |||:::|||   :::|||
      87 GluSerGluSerAlaProGlyAspPheSerLeuSerValLysPheGlyAs 103

    2227 ...............GAAGAATGTCGCCAAGCCACTGGGCGTTGACACA 2194
                             :::      |||:::::::::|||:::
     103 nAspValGlnHisPheLysValLeuArgAspGlyAlaGlyLys..... 117

    2193 GGGTATCCTCCACTCTTTGGCTT......TTTCATACTTTAAACCAGTTG 2150
                        ||||||:::      |||::::::||||||:::|||
     118 ........TyrPheLeuTrpValValLysPheAsnSerLeuAsnGluLeu 131

    2149 GTTCTTTACAGATGAGGACTGTGTTGCTGCGGCATAGATAACCCGTAT.. 2102
         |||   |||:::   :::                      ::::::   :::
     132 ValAspTyrHisArgSerThrSerValSerArgAsnGlnGlnIlePheLe 148

    2101 .......ATTTGGCACCTGCCAAATAAGCCATTAATTTTA.......... 2069
                 |||   :::::::|||::::::|||      :::
     148 uArgAspIleGluGlnValProGlnGlnProThrTyrValGlnAlaLeuP 165

    2068 .........................................GGTCAT 2063
                                                   |||
```

```
 165 heAspPheAspProGlnGluAspGlyGluLeuGlyPheArgArgGlyAsp 181

2062 CTCTGTCACTATCAACAAATCCAGTCACAGAAATAATATGCTGTGAACAT 2013
     :::   ||| :::       :::|||      |||
 182 PheIleHis.ValMetAspAsnSerAspProAsn..............  192

2012 GGCTTTCCTCCTGGTGGGAAGGC......CACTGGGAAGTGAAGGGCTCG 1969
            ||||||:::|||      |||   ::::::
 193 ..........TrpTrpLysGlyAlaCysHisGlyGlnThrGlyMetPhe 205

1968 GTGCGGCGGTACCATTTTTTTTTTTTTTTAAGACTGTGTTTAACCAGTGTG 1919
     |||:::|||                        |||   ||||||
 206 ProArgAsnTyr.......................ValThrProValAs 214

1918 CAGTAAC.....................ACATCTCTTTCTTTCTCTTATT 1890
     |||                        ::::::   :::   :.:   |
 214 nArgAsnValCysAlaAlaAlaAlaGlyAlaAlaThrThrAlaAlaAlaC 231

1889 GCCTGTGCATACGCGCTGCTGACTTGACTCTCACAGAGAAGGTGCGTGCA 1840
     ||   |||      ||||||      |||                 :::
 231 ysCysCysAlaCysAlaAlaGlyCysThr.............GlyCys 242

1839 TCGACTCGTGAAGGTGGGGTCAACAGTGCCCCCATGTGCCTGGATTATCC 1790
     |||      |||:::      :::::::|||   |||
 243 CysThrCysThrGlyAlaCysAlaGlyCysAlaGlyCysCysThrGlyTh 259

1789 TTTTCCAGGT.........GGCCAGCAGTTGCTTATCAGACATCTGCTCT 1749
           |||          ||| :::::::|||   :::      |||:::
 259 rGlyAlaGlyGlyGlyAlaGly.ThrGlyCysAlaGlyAlaAlaCysAla 275

1748 ......GGATAATCCGCAATTGCAAACACACATCCCAATAAGAAGCCTTC 1705
           |||   ::::::   ::::::|||      :::        ::
 276 CysCysGlyThrThrThrThrCysThrThrAlaAlaThrThrThrGlyGl 292

1704 TTCTGGAATCTCCACTGCTGGATCATGTCCAAAAAGCTGATGCTGCTGAG 1655
     ::::      :::|||:::|||:::|||      :::   :::|||   |
 292 yGlyThrGlyThrThrCysGlyAlaCysGlyGlyAlaGlyAlaCysThrG 309

1654 GCGAT............GGTGGTGGCTGCTGTTGCTGCTGCTGCTGC... 1620
     ||              |||||||:::|||:::|||   |||||||||
 309 lyThrCysGlyThrCysGlyGlyAlaCysAlaCysThrCysCysCysThr 325

1619 ..............TGTGAAAATGGATGCGGCGGC............TG 1597
                    |||   :::|||   ||||||            ||
 326 CysAlaCysGlyThrCysThrThrGlyThrGlyGlyThrGlyGlyCysCy 342

1596 CTGGGGGTAAGGTTGCTGAGAGATCTGCTGCTGC..............T 1562
     |   |||      |||      |||||||||                 |
 342 sGlyGlyGlyThrCys......AlaCysCysCysThrGlyThrGlyAlaC 357

1561 GCTGCTGCTGGAGCTGCATTATCTGC....................... 1536
     ||||||||          |||
 357 ysCysCysThrCysThrCysAlaCysThrThrThrGlyGlyThrThrGly 373

1536 ..................................................  1536

 374 GlyAlaAlaCysThrThrThrAlaGlyGlyGlyGlyGlyThrGlyGlyGl 390

1535 ............TGGGGCTGAAGGTGTAAAACCGGGTGCTGC....... 1506
                 |||      |||      |||||||||
 390 yAlaGlyGlyGlyGlyCysAlaCysCysGlyGlyCysCysCysAlaG 407

1506 ..................................................  1506

 407 lyThrGlyGlyGlyAlaCysAlaCysThrGlyGlyGlyAlaGlyAlaGly 423

1506 ..................................................  1506

 424 ThrGlyAlaAlaAlaCysCysAlaAlaCysCysThrThrGlyAlaAlaAl 440

1505 ....TGCTGCTGCTGCTGGGCCTGCTGCTGCTGCTGTAGCATGTGTGTCT 1460
         ||||||||||||   ||||||||||||   |||      |||   :
 440 aThrCysCysCysCysCysAlaCysCysCysThrCysCysCysCysCysG 457
```

```
1459 CTGGAGTCACTTTCACTTGGCTAAACAGCACTGCATTGGCATTTGTATGT 1410
     ::|||:::|||            ::::::||||||   |||   :::
 457 lyGlyThrThr...GlyGlyAlaThrThrThrAlaAlaAlaAlaAlaThr 472

1409 CCCTGC 1404
       |||
 473 GlyCys 474
```

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-415-751-35

```
seq_documentation_block:
; Sequence 35, Application US/08415751
; Patent No. 5643772
;  GENERAL INFORMATION:
;    APPLICANT:  PETERSEN, CAROLYN
;    APPLICANT:  LEECH, JAMES
;    APPLICANT:  NELSON, RICHARD, C.
;    APPLICANT:  GUT, JIRI
;    TITLE OF INVENTION:  POLYPEPTIDES BINDING ANTI-
;    TITLE OF INVENTION:  CRYPTOSPORIDIUM ANTIBODIES, DNA
;    TITLE OF INVENTION:  AND RNA ENCODING THEM, HYBRID
;    TITLE OF INVENTION:  VECTOR AND TRANSFORMED HOST AND
;    TITLE OF INVENTION:  METHODS FOR IMMUNOTHERAPY AND
;    TITLE OF INVENTION:  DIAGNOSIS AND KIT
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PHILLIPS, MOORE, LEMPIO & FINLEY
;      STREET:  385 Sherman Avenue, Suite 6
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94306-1840
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette - 3.5 inch, 1.44 Kb storage
;      COMPUTER:  PC
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  Wordperfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/415,751
;      FILING DATE:  03-APR-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/071,880
;      FILING DATE:  June 1, 1993
;      APPLICATION NUMBER:  07/891,301
;      FILING DATE:  May 29, 1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hana Dolezalova
;      REGISTRATION NUMBER:  30,518
;      REFERENCE/DOCKET NUMBER:  480.19-2 (HHD)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 324-1677
;      TELEFAX:  (415) 324-1678
;  INFORMATION FOR SEQ ID NO:  35:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  362 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    ORIGINAL SOURCE:
;      ORGANISM:  Cryptosporidium parvum
;    FEATURE:
;      NAME/KEY:  Positions coded by nonsense codons are
;      NAME/KEY:  identified as Xaa.
US-08-415-751-35


alignment_scores:
            Quality:  194.00              Length:    232
              Ratio:  2.337                Gaps:     13
Percent Similarity:  35.776   Percent Identity:  22.845
```

```
alignment_block:
 US-09-664-641-10/rev x US-08-415-751-35  ..

 Align seg 1/1  to: US-08-415-751-35  from: 1  to: 362

    1664 TGCTGCTGAGGCGATGGTGGTGGCTGCTGT..................... 1635
          ||||||               |||||||||
     115 CysCys.............GlyCysCysSerLysIleTrpAspAsnCy 126

    1634 ..................TGCTGCTGCTGCTGCTGTGAAAATGGATGCG 1604
                           ||||||||||||||||||   |||
     126 sCysCysSerLeuPheTrpCysCysCysCysCysCysLeuAsnTrpThr. 142

    1603 GCGGCTGCTGGGGGTAAGGTTGCTGAGAGATCTGCTGCTGCTGCTGCTGC 1554
              |||      ||||||         ||||||||||||
     143 .....LeuTrpTrpTyrGlyCys.........CysCysCysCys***Leu 154

    1553 TGG...AGCTGCATTATCTGCTGGGGCTGAAGGTGTAAAACCGGGTGCTG 1507
         |||   :::|||                                |||||
     155 TrpTyrAsnCys..............................CysCy 160

    1506 CTGCTGCTGCTGCTGGGCCTGCTGCTGCTGCTGTAGCATGTGTGTCTCTG 1457
         |||||||   |||   :::||||||||||||||||
     160 sCysCysPheCysCysCysCysCysCysCysHis.............. 172

    1456 GAGTCACTTTCACTTGG...CTAAACAGCACTGCATTGGCATTT...... 1416
                        |||      :::|||:::::::|||:::
     173 ..............TrpArgCysSerSerSerGlyLeuSerValPheHis 184

    1415 ...GTATGTCCCTGCTGG..............CTGTGATTCACCTGCTG 1384
              |||   ||||||                 :::   |||:::|||
     185 CysCysCysArgCysTrp***ArgIleAspThrValAsnPheAsnCysTr 201

    1383 TTCTAAATTTTTCGTAGGTGCTGAAAGAGTCTGT................ 1350
          :::      |||:::           ||||||
     201 pAspTrpAsnPheLeuHisGlnIleLeuValCysSerHisAsnIleValI 218

    1349 ..............................................AAG 1347
                                                      :::
     218 leValThrThr*********ValArgIleIlePheTrpAsnTrpThrArg 234

    1346 ATATGTGCTACATTTGATGGCCGGTTCATCTGCTGAATGTCAGCATTATT 1297
         :::|||:::   ::::::         ||||||   ::::::
     235 LeuCysSerLeuTyrSer.........IleCysAlaIle***GluTrpTh 248

    1296 AGTAATATTCCTCAGTGTC............................. 1278
             :::   |||   |||
     248 r***Val***LeuLeuValProLysGlyIleSerGlyIleThrProArgG 265

    1278 ..............................  ..................  1278

     265 lyAsnSerGlyPheAspProLeuLysThrCy:GlyTyrValLeu***Val 281

    1277 .......................CGTACACCTGGACTCCAGGTAGCTAT 1252
                               |||:::|||   ::::::|||:::
     282 IleAspProLysAspThrAspTyrArgAlaAlaLeuPheLysValThrSe 298

    1251 CATTTCTGATCTTTCAGAACTTTGGAGGTTTTGTCCAGCAGCCATT 1206
          :::|||::::::         |||:::::::|||   :::::::::
     298 rValSerServalValThr***TrpAsnIleCysGluSerThrVal 313

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-07-814-964-13

seq_documentation_block:
; Sequence 13, Application US/07814964
; Patent No. 5359047
;  GENERAL INFORMATION:
;    APPLICANT:  Donahue, Brian A.
;    APPLICANT:  Toney, Jeffrey H.
;    APPLICANT:  Bruhn, Suzanne L.
;    APPLICANT:  Pil, Pieter M.
;    APPLICANT:  Brown, Steven
;    APPLICANT:  Kellett, Patti
;    APPLICANT:  Essigmann, John M.
;    APPLICANT:  Lippard, Stephen J.
;    TITLE OF INVENTION:  DNA Structure Specific Recognition
;    TITLE OF INVENTION:  Protein and Uses Therefor
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  2 Militia Drive
;      CITY:  Lexington
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/814,964
;      FILING DATE:  19911226
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/539,906
;      FILING DATE:  18-JUN-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Granahan, Patricia
;      REGISTRATION NUMBER:  32,227
;      REFERENCE/DOCKET NUMBER:  MIT-4787AAA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-861-6240
;      TELEFAX:  617-861-9540
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  542 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    ORIGINAL SOURCE:
;      ORGANISM:  Saccharomyces cerevisiae
;    IMMEDIATE SOURCE:
;      CLONE:  fractional yeast SSRP (fySSRP) (predicted)
US-07-814-964-13


alignment_scores:
            Quality:  184.00              Length:    509
              Ratio:  0.786                Gaps:     23
 Percent Similarity:  45.972   Percent Identity: 20.236

alignment_block:
 US-09-664-641-10 x US-07-814-964-13  ..

 Align seg 1/1  to: US-07-814-964-13  from: 1  to: 542

     832 TATGAAGAGGAAGAAGAGGAAGAGGAAGAGGAGGAGGAAGTAGAAAATGA 881
           ::::::   :::   ::::::::::::::::::::::::   ::::::::
       4 PheGlnAlaGlnProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGl 20

     882 GGAACAAGATTCTCAGAATGAGGGTAGTACAGATGACAAGTCAAGCCCTG 931
         :::::|||:::::::      :::|||      ::::::
      20 nGlnGlnGlnAlaProTyrGlnGlyHisPheGlnGln.............. 32

     932 CCAGCTCTCAAGAAGGGTCTCCTTCAGGTGACCAGCAG............. 969
                       |||||||:::   :::||||||
      33 ................SerProGlnGlnGlnGlnAsnValTyrPhe 43

     970 ...TTTTCACCTAAATCCAACACTGAAAAATCTAAAGGGGAATTAATGTT 1016
           :::   |||:::|||   |||:::   ::::::   :::   :::
      44 ProLeuProProGlnSerLeuThrGlnProThrSerGlnSerGlnGlnGl 60

    1017 TGATGATTCTTCAGATTCATCACCGGAAAAACAGGAGAGAAATTTAAACT 1066
         :::::::::      :::|||   ::::::::::::::|||:::|||
```

```
  60 nGlnGlnGlnTyrAlaAsnSerAsnSerAsnSerAsnAsnAsnValAsnV 77

1067 GGACCCCGGCCGAAGTCCCACAGTTAGCTGCAGCAAAACGCAGGCTGCCT 1116
       :::         :::|||||||      :::      ::::::
  77 alAsn......AlaLeuProGlnAspPheGlyTyrMetGlnGlnThrGly 91

1117 CAGGGAAAGGAGCCTGGGTTGATTAAC..................... 1143
     :::|||::::::        ||||||
  92 SerGlyGlnAsnTyrProThrIleAsnGlnGlnGlnPheSerGluPheTy 108

1144 .......TTGTGTGCCAATGTCCCACCCGTCCCAGGTAACATTTTGCCCC 1186
             :::   ::::::::              |||         |
 108 rAsnSerPheLeuSerHisLeuThrGlnLysGlnThrAsn.........P 122

1187 CTGAGGTCCGGGGT.............................AAT 1203
     ||:::|||   |||                            |||
 122 roSerValThrGlyThrGlyAlaSerSerAsnAsnAsnSerAsnAsnAsn 138

1204 TTAATGGCTGCTGGACAAAACCTCCAAAGTTCTGAAAGATCAGAAATGAT 1253
       ::::::::::|||:::|||      |||||||:::   :::::::::
 139 AsnValSerSerGlyAsnAsnSerThrSerSerAsnProThrGlnLeuAl 155

1254 AGCTACC...TGGAGTCCAGCTGTACGGACACTGAGGAATATTACTAATA 1300
       |||:::      :::|||||||:::   |||           :::|||
 155 aAlaSerGlnLeuAsnProAlaThrAlaThrThrAlaAlaAlaAsnAsnA 172

1301 ATGCT.....................GACATTCAGCAGATGAACCGG... 1326
       |||                   ::::::   |||:::::::|||
 172 laAlaGlyProAlaSerTyrLeuSerGlnLeuProGlnValGlnArgTyr 188

1327 ...CCATCAAATGTAGCACATATCTTACAGACTCTTTCAGCACCTACGAA 1373
        |||:::|||:::   :::   |||:::::;|||      |||:::::
 189 TyrProAsnAsnMet...AsnAlaLeuSerSerLeuLeuAspProSerSe 204

1374 A.......................................................... 1374
     :
 204 rAlaGlyAsnAlaAlaGlyAsnAlaAsnThrAlaThrHisProGlyLeuL 221

1375 ........AATTTAGAACAGCAGGTGAATCACAGCCAGCAGGGACATACA 1416
             ||||||:::   |||:::::::|||   ||||||
 221 euProProAsnLeuGlnProGlnLeuThrHisHisGlnGln......... 234

1417 AATGCCAATGCAGTGCTGTTTAGCCAAGTGAAAGTGACTCCAGAGACACA 1466

 234 .......................................................... 234

1467 CATGCTACAGCAGCAGCAGGCCCAGCAGCAGCAGCAGCACCCGG 1516
       ::::::|||||||||   |||   ||||||||||||||||
 235 .GlnMetGlnGlnGlnLeuGlnLeuGlnGlnGlnGlnGln..... 247

1517 TTTTACACCTTCAGCCCCAGCAGATAATGCAGCTCCAGCAGCAGCAGCAG 1566
           ||||||   ||||||             |||   |||||||||
 248 ........LeuGlnGlnGlnGln............GlnLeuGlnGlnGln 257

1567 CAGCAGATCTCTCAGCAACCTTACCCCCAGCAGCCGCCGCATCCATTTTC 1616
     :::|||::::::||||||          ||||||      |||   :::::
 258 HisGlnLeuGlnGlnGlnGlnGlnLeuGlnGlnGlnHisHisHisLeuGl 274

1617 ACAGCAGCAGCAGCAGCAACAGCAGCCA..................... 1644
     :|||||||||||||||||||||:::|||
 274 nGlnGlnGlnGlnGlnGlnGlnHisProValValLysLysLeuSerSerT 291

1645 ..........................................CCACCA 1650
                                                 |||
 291 hrGlnSerArgIleGluArgArgLysGlnLeuLysLysGlnGlyProLys 307

1651 TCGCCTCAG.............................CAGCATCAGCT 1670
       |||:::                             :::::::::||
 308 ArgProSerSerAlaTyrPheLeuPheSerMetSerIleArgAsnGluLe 324

1671 TTTTGGACATGATCCAGCAGTGGAGATTCCAGAAGAAGGCTTCTTATTGG 1720
     |:::  :::  |||  :::::::::||||||  :::  |||
 324 uLeuGlnGlnProGluAlaLysValProGluLeuSerLysLeu.... 339

1721 GATGTGTGTTTGCAATTGCGGATTATCCAGAGCAGATGTCTGAT..... 1764
                |||  |||  :::  |||      :::|||
 340 ...........AlaSerAlaArgTrpLysGluLeuThrAspAspGlnLys 352

1765 ...........AAGCAACTGCTGGCCACCTGGAAAAGG.........AT 1793
                 :::::::::  ::::::|||::::::        ::
 353 LysProPheTyrGluGluPheArgThrAsnTrpGluLysTyrArgValVa 369

1794 AATCCAGGCACATGGCGGCACTGTTGACCCC................... 1824
     :  :::|||:::      |||:::  |||
 369 lArgAspAlaTyrGluLysThrLeuProProLysArgProSerGlyProP 386

1825 .....ACCTTCACCAGTCGATGCACGCACCTTCTCTGTGAGAGT..... 1863
           ||||||::::::          :::  ::::::
 386 heIleGlnPheThrGlnGluIleArgProThrValValLysGluAsnPro 402

1864 ..............CAAGTCAGCAGCGCGTATGCACAGGCAATAAGAGA 1898
                  ::::::::::::     ::::::     |||||
 403 AspLysGlyLeuIleGluIleThrLysIleIleGlyGluArgTrpArgGl 419

1899 AAGAAAGAGATGTGTTACTGCACACTGGTTAAACACAGTCTTAAAAAAAA 1948
     |         :::     |||::::::   :::|||     |||:::
 419 uLeuAspProAlaLysLysAlaGluTyrThrGluThrTyrLysLysArgL 436

1949 AAAAAATGGTACCGCCGCACCGAGCCCTTCACTTCCCAGTGGCCTTCCCA 1998
       |||                          ::::::     :::::::|||
 436 euLys.......................GluTrpGluSerCysTyrPro 444

1999 ............CCAGGAGGAAAGCCA 2013
                 |||:::|||:::|||
 445 AspGluAsnAspProAsnGlyAsnPro 453
```

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-258-442-13

```
seq_documentation_block:
; Sequence 13, Application US/08258442
; Patent No. 5670621
;  GENERAL INFORMATION:
;    APPLICANT:  Donahue, Brian A.
;    APPLICANT:  Toney, Jeffrey H.
;    APPLICANT:  Bruhn, Suzanne L.
;    APPLICANT:  Pil, Pieter M.
;    APPLICANT:  Brown, Steven
;    APPLICANT:  Kellett, Patti
;    APPLICANT:  Essigmann, John M.
;    APPLICANT:  Lippard, Stephen J.
;    TITLE OF INVENTION:  DNA Structure Specific Recognition
;    TITLE OF INVENTION:  Protein and Uses Therefor
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  2 Militia Drive
;      CITY:  Lexington
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/258,442
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/539,906
;      FILING DATE:  18-JUN-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Granahan, Patricia
;      REGISTRATION NUMBER:  32,227
;      REFERENCE/DOCKET NUMBER:  MIT-4787AAA
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-861-6240
;      TELEFAX:  617-861-9540
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  542 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    ORIGINAL SOURCE:
;      ORGANISM:  Saccharomyces cerevisiae
;    IMMEDIATE SOURCE:
;      CLONE:  fractional yeast SSRP (fySSRP) (predicted)
US-08-258-442-13


alignment_scores:
            Quality:  184.00              Length:    509
              Ratio:  0.786                Gaps:     23
 Percent Similarity:  45.972   Percent Identity: 20.236

alignment_block:
 US-09-664-641-10 x US-08-258-442-13   ..

 Align seg 1/1  to: US-08-258-442-13  from: 1  to: 542

     832 TATGAAGAGGAAGAAGAGGAAGAGGAAGAGGAGGAGGAAGTAGAAAATGA 881
           ::::::   :::   :::::::::::::::::::::::::   ::::::::
       4 PheGlnAlaGlnProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGl 20

     882 GCAACAAGATTCTCAGAATGAGGGTAGTACAGATGAGAAGTCAAGCCCTG 931
           ::::|||::::::      :::|||      ::::::
      20 nGlnGlnGlnAlaProTyrGlnGlyHisPheGlnGln..... 32

     932 CCAGCTCTCAAGAAGGGTCTCCTTCAGGTGACCAGCAG..... 969
                           |||||||:::   :::||||||
      33 .................SerProGlnGlnGlnGlnGlnAsnValTyrPhe 43

     970 ...TTTTCACCTAAATCCAACACTGAAAAATCTAAAGGGGAATTAATGTT 1016
            :::   |||:::|||   |||:::   ::::::   :::   :::
      44 ProLeuProProGlnSerLeuThrGlnProThrSerGlnSerGlnGlnGl 60

    1017 TGATGATTCTTCAGATTCATCACCGGAAAAACAGGAGAGAAATTTAAACT 1066
           :::::::::      :::|||   ::::::::::::::|||:::|||
      60 nGlnGlnGlnTyrAlaAsnSerAsnSerAsnSerAsnAsnAsnValAsnV 77

    1067 GGACCCCGGCCGAAGTCCCACAGTTAGCTGCAGCAAAACGCAGGCTGCCT 1116
           :::         :::||||||      :::      ::::::
      77 alAsn......AlaLeuProGlnAspPheGlyTyrMetGlnGlnThrGly 91

    1117 CAGGGAAAGGAGCCTGGGTTGATTAAC..... 1143
           :::|||:::::::          ||||||
      92 SerGlyGlnAsnTyrProThrIleAsnGlnGlnGlnPheSerGluPheTy 108

    1144 .......TTGTGTGCCAATGTCCCACCCGTCCCAGGTAACATTTTGCCCC 1186
                :::   :::::::::                  |||         |
     108 rAsnSerPheLeuSerHisLeuThrGlnLysGlnThrAsn.........P 122

    1187 CTGAGGTCCGGGGT.............................AAT 1203
           ||:::|||   |||                           |||
     122 roSerValThrGlyThrGlyAlaSerSerAsnAsnAsnSerAsnAsnAsn 138

    1204 TTAATGGCTGCTGGACAAAACCTCCAAAGTTCTGAAAGATCAGAAATGAT 1253
           :::::::::|||:::|||      ||||||||:::   :::::::::
     139 AsnValSerSerGlyAsnAsnSerThrSerSerAsnProThrGlnLeuAl 155

    1254 AGCTACC...TGGAGTCCAGCTGTACGGACACTGAGGAATATTACTAATA 1300
           |||:::      :::|||||||:::   |||             :::|||
     155 aAlaSerGlnLeuAsnProAlaThrAlaThrThrAlaAlaAlaAsnAsnA 172

    1301 ATGCT.....................GACATTCAGCAGATGAACCGG... 1326
           |||                     ::::::   |||::::::|||
     172 laAlaGlyProAlaSerTyrLeuSerGlnLeuProGlnValGlnArgTyr 188

    1327 ...CCATCAAATGTAGCACATATCTTACAGACTCTTTCAGCACCTACGAA 1373
            |||:::|||:::   :::   |||:::::::|||      |||:::::
     189 TyrProAsnAsnMet...AsnAlaLeuSerSerLeuLeuAspProSerSe 204

    1374 A..... 1374
            :
     204 rAlaGlyAsnAlaAlaGlyAsnAlaAsnThrAlaThrHisProGlyLeuL 221

    1375 ........AATTTAGAACAGCAGGTGAATCACAGCCAGCAGGGACATACA 1416
                 ||||||:::   |||:::::::|||   ||||||
     221 euProProAsnLeuGlnProGlnLeuThrHisHisGlnGln..... 234

    1417 AATGCCAATGCAGTGCTGTTTAGCCAAGTGAAAGTGACTCCAGAGACACA 1466

     234 ..... 234

    1467 CATGCTACAGCAGCAGCAGCAGGCCCAGCAGCAGCAGCAGCAGCACCCGG 1516
          ::::::||||||||||   |||   ||||||||||||||||
     235 .GlnMetGlnGlnGlnLeuGlnLeuGlnGlnGlnGlnGln..... 247

    1517 TTTTACACCTTCAGCCCCAGCAGATAATGCAGCTCCAGCAGCAGCAGCAG 1566
                |||||||   ||||||              |||   |||||||||
     248 ........LeuGlnGlnGlnGln............GlnLeuGlnGlnGln 257

    1567 CAGCAGATCTCTCAGCAACCTTACCCCCAGCAGCCGCCGCATCCATTTTC 1616
          :::|||::::::||||||          ||||||      |||   :::::
     258 HisGlnLeuGlnGlnGlnGlnGlnLeuGlnGlnGlnHisHisHisLeuGl 274

    1617 ACAGCAGCAGCAGCAGCAACAGCAGCCA..... 1644
          :|||||||||||||||||||||||:::|||
     274 nGlnGlnGlnGlnGlnGlnGlnHisProValValLysLysLeuSerSerT 291

    1645 .....CCACCA 1650
                                        |||
     291 hrGlnSerArgIleGluArgArgLysGlnLeuLysLysGlnGlyProLys 307

    1651 TCGCCTCAG.............................CAGCATCAGCT 1670
            |||:::                           ::::::::::||
     308 ArgProSerSerAlaTyrPheLeuPheSerMetSerIleArgAsnGluLe 324

    1671 TTTTGGACATGATCCAGCAGTGGAGATTCCAGAAGAAGGCTTCTTATTGG 1720
          |:::   :::   |||   ::::::::||||||   :::   |||
     324 uLeuGlnGlnProGluAlaLysValProGluLeuSerLysLeu.... 339

    1721 GATGTGTGTTTGCAATTGCGGATTATCCAGAGCAGATGTCTGAT..... 1764
                  |||   |||   :::   |||      ::|||
     340 ..........AlaSerAlaArgTrpLysGluLeuThrAspAspGlnLys 352

    1765 ............AAGCAACTGCTGGCCACCTGGAAAAGG.........AT 1793
                      :::::::::   ::::::|||::::::         ::
     353 LysProPheTyrGluGluPheArgThrAsnTrpGluLysTyrArgValVa 369

    1794 AATCCAGGCACATGGCGGCACTGTTGACCCC..... 1824
          :   :::|||:::      |||:::   |||
     369 lArgAspAlaTyrGluLysThrLeuProProLysArgProSerGlyProP 386

    1825 .....ACCTTCACGAGTCGATGCACGCACCTTCTCTGTGAGAGT..... 1863
                |||||||::::::              :::   ::::::
     386 heIleGlnPheThrGlnGluIleArgProThrValValLysGluAsnPro 402

    1864 ..............CAAGTCAGCAGCGCGTATGCACAGGCAATAAGAGA 1898
                        :::::::::::      ::::::      |||||
     403 AspLysGlyLeuIleGluIleThrLysIleIleGlyGluArgTrpArgGl 419

    1899 AAGAAAGAGATGTGTTACTGCACACTGGTTAAACACAGTCTTAAAAAAAA 1948
          |         :::      |||::::::   :::|||      |||:::
     419 uLeuAspProAlaLysLysAlaGluTyrThrGluThrTyrLysLysArgL 436

    1949 AAAAAATGGTACCGCCGCACCGAGCCCTTCACTTCCCAGTGGCCTTCCCA 1998
          |||                          ::::::      ::::::|||
     436 euLys.....................GluTrpGluSerCysTyrPro 444
```

```
 1999 ...........CCAGGAGGAAAGCCA 2013
                  |||:::|||:::|||
  445 AspGluAsnAspProAsnGlyAsnPro 453

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-328-809-8

seq_documentation_block:
; Sequence 8, Application US/08328809
; Patent No. 5705334
;  GENERAL INFORMATION:
;    APPLICANT:  Lippard, Stephen J.
;    APPLICANT:  Essigmann, John M.
;    APPLICANT:  Donahue, Brian A.
;    APPLICANT:  Toney, Jeffrey H.
;    APPLICANT:  Bruhn, Suzanne L.
;    APPLICANT:  Pil, Pieter M.
;    APPLICANT:  Brown, Steven
;    APPLICANT:  Kellett, Patti
;    TITLE OF INVENTION:  Uses For DNA Structure-Specific
;    TITLE OF INVENTION:  Recognition Proteins
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Patent Administrator, Testa, Hurwitz & Thibeault
;      STREET:  53 State Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/328,809
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fenton, Gillian M.
;      REGISTRATION NUMBER:  36,508
;      REFERENCE/DOCKET NUMBER:  MIT-023 (5473/24)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-248-7000
;      TELEFAX:  617-248-7100
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  542 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    ORIGINAL SOURCE:
;      ORGANISM:  Saccharomyces cerevisiae
;    IMMEDIATE SOURCE:
;      CLONE:  fractional yeast SSRP (fySSRP) (predicted)
US-08-328-809-8


alignment_scores:
            Quality:  184.00              Length:    509
              Ratio:  0.786                Gaps:     23
Percent Similarity:  45.972   Percent Identity: 20.236

alignment_block:
US-09-664-641-10 x US-08-328-809-8   ..

Align seg 1/1  to: US-08-328-809-8  from: 1  to: 542

  832 TATGAAGAGGAAGAAGAGGAAGACGAAGAGGAGGAGGAAGTAGAAAATGA 881
      ::::::   :::   ::::::::::::::::::::::::   ::::::::
    4 PheGlnAlaGlnProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGl 20

  882 GGAACAAGATTCTCAGAATGAGGGTACTACAGATGAGAAGTCAAGCCCTG 931
      ::::|||::::::      :::|||      ::::::
   20 nGlnGlnGlnAlaProTyrGlnGlyHisPheGlnGln..... 32

  932 CCAGCTCTCAAGAAGGGTCTCCTTCAGGTGACCAGCAG..... 969
                    |||||||:::   :::|||||
   33 ................SerProGlnGlnGlnGlnGlnAsnValTyrPhe 43

  970 ...TTTTCACCTAAATCCAACACTGAAAAATCTAAAGGGGAATTAATGTT 1016
        :::   |||:::|||   |||:::   ::::::   :::   :::
   44 ProLeuProProGlnSerLeuThrGlnProThrSerGlnSerGlnGlnGl 60

 1017 TGATGATTCTTCAGATTCATCACCGGAAAAACAGGAGAGAAATTTAAACT 1066
        :::::::::      :::|||   ::::::::::::::::|||:::|||
   60 nGlnGlnGlnTyrAlaAsnSerAsnSerAsnSerAsnAsnAsnValAsnV 77

 1067 GGACCCCGGCCGAAGTCCCACAGTTAGCTGCAGCAAAACGCAGGCTGCCT 1116
        :::         :::|||||||      :::      ::::::
   77 alAsn......AlaLeuProGlnAspPheGlyTyrMetGlnGlnThrGly 91

 1117 CAGGGAAAGGAGCCTGGGTTGATTAAC..... 1143
      :::|||::::::         ||||||
   92 SerGlyGlnAsnTyrProThrIleAsnGlnGlnGlnPheSerGluPheTy 108

 1144 .......TTGTGTGCCAATGTCCCACCCGTCCCAGGTAACATTTTGCCCC 1186
              :::   :::::::::                |||         |
  108 rAsnSerPheLeuSerHisLeuThrGlnLysGlnThrAsn.........P 122

 1187 CTGAGGTCCGGGGT.................................AAT 1203
      ||:::|||   |||                                  |||
  122 roSerValThrGlyThrGlyAlaSerSerAsnAsnAsnSerAsnAsnAsn 138

 1204 TTAATGGCTGCTGGACAAAACCTCCAAAGTTCTGAAAGATCAGAAATGAT 1253
        :::::::::|||:::|||      |||||||:::   :::::::::
  139 AsnValSerSerGlyAsnAsnSerThrSerSerAsnProThrGlnLeuAl 155

 1254 AGCTACC...TGGAGTCCAGCTGTACGGACACTGAGGAATATTACTAATA 1300
       |||:::      :::|||||||:::   |||             :::|||
  155 aAlaSerGlnLeuAsnProAlaThrAlaThrThrAlaAlaAlaAsnAsnA 172

 1301 ATGCT.....................GACATTCAGCAGATGAACCGG... 1326
        |||                     ::::::   |||::::::|||
  172 laAlaGlyProAlaSerTyrLeuSerGlnLeuProGlnValGlnArgTyr 188

 1327 ...CCATCAAATGTAGCACATATCTTACAGACTCTTTCAGCACCTACGAA 1373
        |||:::|||:::   :::   |||::::::|||      |||:::::
  189 TyrProAsnAsnMet...AsnAlaLeuSerSerLeuLeuAspProSerSe 204

 1374 A..................................................... 1374
      :
  204 rAlaGlyAsnAlaAlaGlyAsnAlaAsnThrAlaThrHisProGlyLeuL 221

 1375 ........AATTTAGAACAGCAGGTGAATCACAGCCAGCAGGGACATACA 1416
              ||||||:::   |||::::::|||   ||||||
  221 euProProAsnLeuGlnProGlnLeuThrHisHisGlnGln......... 234

 1417 AATGCCAATGCAGTGCTGTTTAGCCAAGTGAAAGTGACTCCAGAGACACA 1466

  234 ..................................................... 234

 1467 CATGCTACAGCAGCAGCAGGCCCAGCAGCAGCAGCAGCAGCACCCGG 1516
        ::::::|||||||||   |||   ||||||||||||||
  235 .GlnMetGlnGlnGlnLeuGlnLeuGlnGlnGlnGlnGln......... 247

 1517 TTTTACACCTTCAGCCCCAGCAGATAATGCAGCTCCAGCAGCAGCAGCAG 1566
             ||||||   ||||||             |||   |||||||||
  248 ........LeuGlnGlnGlnGln............GlnLeuGlnGlnGln 257

 1567 CAGCAGATCTCTCAGCAACCTTACCCCCAGCAGCCGCCGCATCCATTTTC 1616
      :::|||::::::||||||              ||||||      |||   :::::
  258 HisGlnLeuGlnGlnGlnGlnGlnLeuGlnGlnGlnHisHisHisLeuGl 274

 1617 ACAGCAGCAGCAGCAGCAACAGCAGCCA..................... 1644
      :|||||||||||||||||||||:::|||
  274 nGlnGlnGlnGlnGlnGlnGlnHisProValValLysLysLeuSerSerT 291
```

```
 1645 .....................................CCACCA 1650
                                             |||
  291 hrGlnSerArgIleGluArgArgLysGlnLeuLysLysGlnGlyProLys 307

 1651 TCGCCTCAG.............................CAGCATCAGCT 1670
        |||:::                              ::::::::||
  308 ArgProSerSerAlaTyrPheLeuPheSerMetSerIleArgAsnGluLe 324

 1671 TTTTGGACATGATCCAGCAGTGGAGATTCCAGAAGAAGGCTTCTTATTGG 1720
      |:::   :::   |||   :::::::::||||||   :::   |||
  324 uLeuGlnGlnPheProGluAlaLysValProGluLeuSerLysLeu.... 339

 1721 GATGTGTGTTTGCAATTGCGGATTATCCAGAGCAGATGTCTGAT...... 1764
               |||   |||   :::   |||       :::|||
  340 ..........AlaSerAlaArgTrpLysGluLeuThrAspAspGlnLys 352

 1765 ............AAGCAACTGCTGGCCACCTGGAAAAGG.... ....AT 1793
                  :::::::::   ::::::|||:::::         ::
  353 LysProPheTyrGluGluPheArgThrAsnTrpGluLysTyrArgValVa 369

 1794 AATCCAGGCACATGGCGGCACTGTTGACCCC..................... 1824
      :   :::|||:::      |||:::   |||
  369 lArgAspAlaTyrGluLysThrLeuProProLysArgProSerGlyProP 386

 1825 .....ACCTTCACGAGTCGATGCACGCACCTTCTCTGTGAGAGT...... 1863
           ||||||::::::            :::   ::::::
  386 heIleGlnPheThrGlnGluIleArgProThrValValLysGluAsnPro 402

 1864 ..............CAAGTCAGCAGCGCGTATGCACAGGCAATAAGAGA 1898
                    ::::::::::::     ::::::      |||||
  403 AspLysGlyLeuIleGluIleThrLysIleIleGlyGluArgTrpArgGl 419

 1899 AAGAAAGAGATGTGTTACTGCACACTGGTTAAACACAGTCTTAAAAAAAA 1948
      |      :::      |||:::::::   :::|||      |||:::
  419 uLeuAspProAlaLysLysAlaGluTyrThrGluThrTyrLysLysArgL 436

 1949 AAAAAATGGTACCGCCGCACCGAGCCCTTCACTTCCCAGTGGCCTTCCCA 1998
        |||                       ::::::      :::::::|||
  436 euLys.....................GluTrpGluSerCysTyrPro 444

 1999 ............CCAGGAGGAAAGCCA 2013
                  |||:::|||:::|||
  445 AspGluAsnAspProAsnGlyAsnPro 453
```

seq_name: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:PCT-US92-11107-13

```
seq_documentation_block:
; Sequence 13, Application PC/TUS9211107
;  GENERAL INFORMATION:
;    APPLICANT:  Donahue, Brian A.
;    APPLICANT:  Toney, Jeffrey H.
;    APPLICANT:  Bruhn, Suzanne L.
;    APPLICANT:  Pil, Pieter M.
;    APPLICANT:  Brown, Steven
;    APPLICANT:  Kellett, Patti
;    APPLICANT:  Essigmann, John M.
;    APPLICANT:  Lippard, Stephen J.
;    TITLE OF INVENTION:  DNA Structure Specific Recognition
;    TITLE OF INVENTION:  Protein and Uses Therefor
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  2 Militia Drive
;      CITY:  Lexington
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/11107
;      FILING DATE:  19921218
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/539,906
;      FILING DATE:  18-JUN-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Granahan, Patricia
;      REGISTRATION NUMBER:  32,227
;      REFERENCE/DOCKET NUMBER:  MIT-4787AAA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-861-6240
;      TELEFAX:  617-861-9540
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  542 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    ORIGINAL SOURCE:
;      ORGANISM:  Saccharomyces cerevisiae
;    IMMEDIATE SOURCE:
;      CLONE:  fractional yeast SSRP (fySSRP) (predicted)
PCT-US92-11107-13


alignment_scores:
            Quality:  184.00              Length:    509
              Ratio:  0.786                Gaps:     23
 Percent Similarity:  45.972   Percent Identity: 20.236

alignment_block:
 US-09-664-641-10 x PCT-US92-11107-13   ..

 Align seg 1/1  to: PCT-US92-11107-13  from: 1  to: 542

  832 TATGAAGAGGAAGAAGAGGAAGAGGAAGAGGAGGAGGAAGTAGAAAATGA 881
      ::::::   :::   :::::::::::::::::::::::   ::::::::
    4 PheGlnAlaGlnProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGl 20

  882 GGAACAAGATTCTCAGAATGAGGGTAGTACAGATGAGAAGTCAAGCCCTG 931
      ::::|||::::::      :::|||      ::::::
   20 nGlnGlnGlnAlaProTyrGlnGlyHisPheGlnGln.............. 32

  932 CCAGCTCTCAAGAAGGGTCTCCTTCAGGTGACCAGCAG............. 969
                    ||||||:::   :::||||||
   33 ................SerProGlnGlnGlnGlnAsnValTyrPhe 43

  970 ...TTTTCACCTAAATCCAACACTGAAAAATCTAAAGGGGAATTAATGTT 1016
        :::   |||:::|||   |||:::   ::::::   :::   :::
   44 ProLeuProProGlnSerLeuThrGlnProThrSerGlnSerGlnGlnGl 60

 1017 TGATGATTCTTCAGATTCATCACCGGAAAAACAGGAGAGAAATTTAAACT 1066
      :::::::::      :::|||   ::::::::::::::::|||:::|||
   60 nGlnGlnGlnTyrAlaAsnSerAsnSerAsnSerAsnAsnAsnValAsnV 77

 1067 GGACCCCGGCCCGAAGTCCCACAGTTAGCTGCAGCAAAACGCAGGCTGCCT 1116
        :::         :::||||||      :::      ::::::
   77 alAsn......AlaLeuProGlnAspPheGlyTyrMetGlnGlnThrGly 91

 1117 CAGGGAAAGGAGCCTGGGTTGATTAAC..................... 1143
      :::|||::::::         ||||||
   92 SerGlyGlnAsnTyrProThrIleAsnGlnGlnGlnPheSerGluPheTy 108

 1144 .......TTGTGTGCCAATGTCCCACCCGTCCCAGGTAACATTTTGCCCC 1186
            :::   :::::::::                |||       |
  108 rAsnSerPheLeuSerHisLeuThrGlnLysGlnThrAsn.........P 122

 1187 CTGAGGTCCGGGGT.................................AAT 1203
      ||:::|||   |||                                 |||
  122 roSerValThrGlyThrGlyAlaSerSerAsnAsnAsnSerAsnAsnAsn 138
```

```
1204 TTAATGGCTGCTGGACAAAACCTCCAAAGTTCTGAAAGATCAGAAATGAT 1253
     ::::::::::|||:::|||      |||||||:::   :::::::::
 139 AsnValSerSerGlyAsnAsnSerThrSerSerAsnProThrGlnLeuAl 155

1254 AGCTACC...TGGAGTCCAGCTGTACGGACACTGAGGAATATTACTAATA 1300
      |||:::      :::|||||||:::   |||           :::|||
 155 aAlaSerGlnLeuAsnProAlaThrAlaThrThrAlaAlaAlaAsnAsnA 172

1301 ATGCT.....................GACATTCAGCAGATGAACCGG... 1326
       |||                    ::::::   |||::::::|||
 172 laAlaGlyProAlaSerTyrLeuSerGlnLeuProGlnValGlnArgTyr 188

1327 ...CCATCAAATGTAGCACATATCTTACAGACTCTTTCAGCACCTACGAA 1373
        |||:::|||:::   :::   |||::::::|||      |||:::::
 189 TyrProAsnAsnMet...AsnAlaLeuSerSerLeuLeuAspProSerSe 204

1374 A.............................................. 1374
     :
 204 rAlaGlyAsnAlaAlaGlyAsnAlaAsnThrAlaThrHisProGlyLeuL 221

1375 ........AATTTAGAACAGCAGGTGAATCACAGCCAGCAGGGACATACA 1416
             |||||||:::   |||::::::|||   ||||||
 221 euProProAsnLeuGlnProGlnLeuThrHisHisGlnGln......... 234

1417 AATGCCAATGCAGTGCTGTTTAGCCAAGTGAAAGTGACTCCAGAGACACA 1466

 234 .............................................. 234

1467 CATGCTACAGCAGCAGCAGCAGGCCCAGCAGCAGCAGCAGCAGCACCCGG 1516
      ::::::|||||||||   |||   ||||||||||||||||
 235 .GlnMetGlnGlnGlnLeuGlnLeuGlnGlnGlnGlnGln.......... 247

1517 TTTTACACCTTCAGCCCCAGCAGATAATGCAGCTCCAGCAGCAGCAGCAG 1566
             ||||||  ||||||            |||   |||||||||
 248 ........LeuGlnGlnGlnGln............GlnLeuGlnGlnGln 257

1567 CAGCAGATCTCTCAGCAACCTTACCCCCAGCAGCCGCCGCATCCATTTTC 1616
     :::|||:::::::|||||         ||||||      |||   :::::
 258 HisGlnLeuGlnGlnGlnGlnGlnLeuGlnGlnGlnHisHisHisLeuGl 274

1617 ACAGCAGCAGCAGCAGCAACAGCAGCCA...................... 1644
     :|||||||||||||||||||:::|||
 274 nGlnGlnGlnGlnGlnGlnGlnHisProValValLysLysLeuSerSerT 291

1645 .....................................CCACCA 1650
                                          |||
 291 hrGlnSerArgIleGluArgArgLysGlnLeuLysLysGlnGlyProLys 307

1651 TCGCCTCAG.............................CAGCATCAGCT 1670
       |||:::                            ::::::::::||
 308 ArgProSerSerAlaTyrPheLeuPheSerMetSerIleArgAsnGluLe 324

1671 TTTTGGACATGATCCAGCAGTGGAGATTCCAGAAGAAGGCTTCTTATTGG 1720
     |:::   :::   |||   :::::::::||||||   :::   |||
 324 uLeuGlnGlnPheProGluAlaLysValProGluLeuSerLysLeu.... 339

1721 GATGTGTGTTTGCAATTGCGGATTATCCAGAGCAGATGTCTGAT...... 1764
               |||   |||   :::   |||      :::|||
 340 ...........AlaSerAlaArgTrpLysGluLeuThrAspAspGlnLys 352

1765 ............AAGCAACTGCTGGCCACCTGGAAAAGG.........AT 1793
                :::::::::   ::::::|||::::::         ::
 353 LysProPheTyrGluGluPheArgThrAsnTrpGluLysTyrArgValVa 369

1794 AATCCAGGCACATGGCGGCACTGTTGACCCC.................... 1824
     :   :::|||:::      |||:::   |||
 369 lArgAspAlaTyrGluLysThrLeuProProLysArgProSerGlyProP 386

1825 .....ACCTTCACGAGTCGATGCACGCACCTTCTCTGTGAGAGT...... 1863
           ||||||::::::          :::   ::::::
 386 heIleGlnPheThrGlnGluIleArgProThrValValLysGluAsnPro 402

1864 ................CAAGTCAGCAGCGCGTATGCACAGGCAATAAGAGA 1898
```

```
                    ::::::::::::      ::::::      |||||
 403 AspLysGlyLeuIleGluIleThrLysIleIleGlyGluArgTrpArgGl 419

1899 AAGAAAGAGATGTGTTACTGCACACTGGTTAAACACAGTCTTAAAAAAAA 1948
     |          :::      |||::::::   :::|||      |||:::
 419 uLeuAspProAlaLysLysAlaGluTyrThrGluThrTyrLysLysArgL 436

1949 AAAAAATGGTACCGCCGCACCGAGCCCTTCACTTCCCAGTGGCCTTCCCA 1998
       |||                          ::::::      ::::::|||
 436 euLys.......................GluTrpGluSerCysTyrPro 444

1999 ............CCAGGAGGAAAGCCA 2013
                 |||:::|||:::|||
 445 AspGluAsnAspProAsnGlyAsnPro 453
```

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-990-823-112

seq_documentation_block:
; Sequence 112, Application US/08990823D
; Patent No. 6228371
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; TITLE OF INVENTION: immunostimulatory Peptides
; FILE REFERENCE: 49086
; CURRENT APPLICATION NUMBER: US/08/990,823D
; CURRENT FILING DATE: 1997-12-15
; EARLIER APPLICATION NUMBER: US 96/10375
; EARLIER FILING DATE: 1996-06-14
; EARLIER APPLICATION NUMBER: 60/000,254
; EARLIER FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 113
; SOFTWARE: PatentIN Ver. 2.0
; SEQ ID NO 112
; LENGTH: 143
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-08-990-823-112

alignment_scores:
Quality: 182.50 Length: 160
Ratio: 2.724 Gaps: 13
Percent Similarity: 41.875 Percent Identity: 31.875

alignment_block:
US-09-664-641-10/rev x US-08-990-823-112 ..

Align seg 1/1 to: US-08-990-823-112 from: 1 to: 143

```
1688 GCTGGATCATGTCCAAAAAGCTGATGCTGCTGAGGCGATGGTGGTGGCTG 1639
     ||||||:::|||            :::|||   |||      ||||||||
    1 AlaGlyGlyCysGly.........AlaCysGlyGlyThrThrGlyGlyCy 14

1638 CTGTTGCTGCTGCTGCTGC.............................T 1619
     |:::|||||||:::||||||                            |
  14 sAlaCysCysAlaCysCysAlaCysGlyAlaGlyThrGlyAlaGlyThrC 31

1618 GTGAAAATGGATGCGGCGGCTGCTGGGGGTAAGGTTGCTGAGAGATCTGC 1569
     ||   :::||||||:::|||      |||   ||||||      |||
  31 ysAlaGlyGlyCysAlaGly......GlyAlaGlyCys.........Cys 42

1568 TGCTGC...TGCTGC................................. 1557
     ||||||   ||||||
  43 CysCysGlyCysCysAlaCysGlyThrThrGlyCysGlyGlyAlaCysGl 59

1556 .......................TGCTGGAGCTGCATTATCTGCTGGG 1532
                            |||   :::|||      |||   |
  59 yGlyCysGlyCysGlyAlaAlaThrCysThrThrCysGlyCysCysThrG 76

1531 GCTGAAGGTGTAAAACCGGGTGCTGCTGCTGCTGCTGCTGGGCCTGC... 1485
     ||                ||||||   |||:::|||||   ||||||
  76 ly.........ThrGlyGlyCysThrCysAlaCysCysGlyAlaCysThr 89
```

```
 1484 ..............TGCTGCTGC......TGTAGCATGTGTGTCTCT.. 1458
                    |||||||||      |||:::  |||::::::
   90 ThrThrCysCysGlyCysCysCysAlaAlaCysThrThrCysAlaAlaCy 106

 1457 .GGAGTCACTTTCACTTGGCTAAACAGCACTGCATTGGCATTTGTATGT. 1410
       |||:::|||  |||         :::  |||  :::  :::|||
  106 sGlyAlaThrCysThrThrGlyCysAlaCysAlaThrGlyGlyAlaCysG 123

 1409 ...................CCCTGCTGGCTGTGATTCACCTGCTGTTCT 1380
                          |||               ||||||
  123 lyGlyCysAlaAlaGlyAlaAlaCys...........GlyCysCys... 134

 1379 AAATTTTTCGTAGGTGCTGAAAGAGTCTGT 1350
               ||||||      :::|||
  135 ...........GlyAlaGlyGlyThrCys 140
```

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-916-352-2

```
seq_documentation_block:
; Sequence 2, Application US/08916352
; Patent No. 6166191
;  GENERAL INFORMATION:
;    APPLICANT:  CHIRON CORPORATION
;    TITLE OF INVENTION:  HUMAN POLYHOMEOTIC 1 (hph1) ACTS AS A
;    TITLE OF INVENTION:  TUMOR SUPPRESSOR
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CHIRON CORPORATION
;      STREET:  4560 HORTON STREET
;      CITY:  EMERYVILLE
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94608
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/916,352
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  POTTER, JANE
;      REGISTRATION NUMBER:  33,332
;      REFERENCE/DOCKET NUMBER:  1355.
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  510-923-2707
;      TELEFAX:  510-655-3542
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1004 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-916-352-2


alignment_scores:
            Quality:  174.00              Length:    693
              Ratio:  0.563                 Gaps:     31
Percent Similarity:  44.589   Percent Identity: 21.934

alignment_block:
US-09-664-641-10 x US-08-916-352-2   ..

Align seq 1/1  to: US-08-916-352-2  from: 1  to: 1004

    736 GCAAGTATTAAAATTGTGACTCCTGACTGGGTTCTGGATTGCGTATCAGA 785
        ||||||      |||:::  |||:::  ::::::  :::|||:::::
    172 AlaSerGlnLeuIleLeuMetProAsnGlyAlaValAlaAlaValGlnGl 188

    786 GAAAACCAAAAAGGACGAAGCATTTTATCATCCTCGTCTGATTATTTATG 835
        :::::::  :::  ::::::      |||  :::
    188 nGluValProSerAlaGlnSer.........ProGlyVal......... 198

    836 AAGAGGAAGAAGAGGAAGAGGAAGAGGAGGAGGAAGTAGAAAATGAGGAA 885
          :::  :::  ::::::  ::::::      |||:::|||::::::
    199 ..HisAlaAspAlaAspGlnValGlnAsnLeuAlaValArgAsnGlnGln 214

    886 .........CAAGATTCTCAGAATGAGGGTAGTACAGATGAGAAGTCAAG 926
                |||      |||  :::||||||||||::::::
    215 AlaSerAlaGlnGlyProGlnMetGlnGlySerThrGlnLysAlaIlePr 231

    927 CCCTGCCAGCTCTCAAGAAGGGTCTCCTTCAGGTGACCAGCAGTTTTCAC 976
        |||::::::                                    ||||
    231 oProGlyAla.....................................SerP 236

    977 CTAAATCCAACACTGAAAAATCTAAAGGGGAATTAATGTTTGATGATTCT 1026
        ||   |||:::  ::::::::::::::::::  :::       :::
    236 roValSerSerLeuSerGlnAlaSerSerGlnAlaLeuAlaValAlaGln 252

   1027 TCAGATTCATCACCGGAAAAACAGGAGAGAAATTTAAACTGGACC..... 1071
        ::::::|||:::      :::|||:::  ||||||:::  :::
    253 AlaSerSerGlyAlaThrAsnGlnSerLeuAsnLeuSerGlnAlaGlyGl 269

   1072 .................................CCGGCCGAAGTCCCAC 1087
                                        |||:::       |
    269 yGlySerGlyAsnSerIleProGlySerMetGlyProGlyGlyGlyG 286

   1088 AGTTAGCTGCAGCAAAACGCAGGCTGCCTCAGGGAAAGGAGCCTGGGTTG 1137
        ||      ::::::      :::||||||::::::        |||
    286 lnAlaHisGlyGlyLeuGlyGlnLeuProSerSerGlyMetGlyGly... 301

   1138 ATTAACTTGTGTGCCAATGTCCCACCCGTCCCAGGTAACATTTTGCCCCC 1187
                :::::  |||                :::::::::  ||
    302 ........ ...GlySerCysProArgLysGlyThrGlyValValGlnPr 314

   1188 TGAGGTCCGGGGTAATTTAATGGCTGCTGGACAAAACCTCCAAAGTTCTG 1237
        |                 :::  |||:::|||::::::      |||:
    314 o.................LeuProAlaAlaGlnThrValThrValSerG 325

   1238 AAAGATCAGAAATGATAGCTACCTGGAGTCCAGCTGTACGGACA...... 1281
        ::  |||:::      |||      :::  |||  ::::::
    325 lnGlySerGlnThrGluAlaGluSerAlaAlaAlaLysLysAlaGluAla 341

   1282 ............CTGAGGAATATTACTAATAATGCTGACATTCAGCAGAT 1319
                    :::|||:::      |||
    342 AspGlySerGlyGlnGlnAsnValGlyMetAsnLeuThrArgThrAlaTh 358

   1320 GAACCGGCCATCAAAT......GTAGCACATATCTTACAGACTCTTTCAG 1363
             |||||||:::      ::::::           |||
    358 rProAlaProSerGlnThrLeuIleSerSerAlaThrTyrThrGlnIleG 375

   1364 CACCTACGAAAAATTTAGAACAGCAGGTGAATCACAGCCAGCAGGGACAT 1413
          |||   :::  ::::::||||||                ||||||  |||
    375 lnProHisSerLeuIleGlnGlnGln............GlnGlnIleHis 387

   1414 ACAAATGCCAATGCAGTGCTGTTTAGCCAAGTGAAAGTGACTCCAGAGAC 1463
          :::  :::  |||::::::::::|||      :::      :::
    388 LeuGlnGlnLysGlnValValIleGlnGlnGlnIleAlaIleHisHis.. 403

   1464 ACACA.GCTACAGCAGCAGCAGCAGGCCCAGCAGCAGCAGCAGCACC 1513
                |||||||||  |||  :::|||:::|||      |||
    404 .........GlnGlnGlnPheGlnHisArgGlnSerGlnLeuLeuHisT 417

   1514 CGGTTTTACACCTTCAGCCCCAGCAGATAATGCAGCTCCAGCAGCAGCAG 1563
          :::  ||||||          |||:::  |||  ||||||||||||
    417 hrAlaThrHisLeu.........GlnLeuAlaGlnGlnGlnGlnGlnGln 430

   1564 CAGCAGCAGATCTCTCAGCAACCTTACCCCCAG.................. 1596
        |||||||||  :::|||||||      ||||||
    431 GlnGlnGlnGlnGlnGlnGlnGlnGlnProGlnAlaThrThrLeuThrAl 447
```

```
1597 ....CAGCCGCCGCATCCATTTTCACAGCAGCAGCAGCAGCAACAGCAGC 1642
          |||||||||              |||             ||||||
 447 aProGlnProPro..............GlnValProProThrGlnGlnV 459

1643 CACCACCATCGCCTCAGCAGCATCAGCTTTTTGGACATGATCCAGCAGTG 1692
       |||||||||   :::|||:::|||          ::::::   |||
 459 alProProSerGlnSerGlnGlnGlnAla.GlnThrLeuValValGlnPr 475

1693 GAGATTCCAGAAGAAGGCTTCTTATTGGGATGTGTGTTTGCAATTGCGG. 1741
       :::|||::::::                    |||:::|||
 475 oMetLeuGlnSerSerPro.................LeuSerLeuProP 486

1742 .........................ATTATCCAGAGCAGATGTCTGATA 1765
                               ||||||||||:::      :::
 486 roAspAlaAlaProLysProProIleProIleGlnSerLysProProVal 502

1766 AGCAACTGCTGGCCACCTGGAAAAGGATAA[illegible]AGGCACATGGCGGCACT 1815
     :::         ||||||     |||   :::      |||:::|||
 503 AlaProIleLysProProGlnLeuCly...AlaAlaLysMetSerAlaAl 518

1816 GTTGACCCCACCTTCACGAGTCGATGCACGCACCTTCTCTGTGAGAGTCA 1865
           ||||||   ::::::   :::::::::   :::::::|||
 518 aGlnGlnProProProHisIleProValGlnValValGlyThrArgGlnP 535

1866 AGTCAGCAGCGCCTATGCACAGGCAATAAGACAAAGAAAGAGATGTGTTA 1915
       ::::::|||:::   :::                :::      :::
 535 roGlyThrAlaGlnAlaGlnAlaLeuGlyLeuAlaGlnLeuAlaAlaAla 551

1916 CTGCACACT..............GGTTAAACACAGTCTTAAAAAAAAAA 1950
     :::   |||              |||   :::||||||    :::
 552 ValProThrSerArgGlyMetProGlyThrValGlnSer.GlyGlnAlaH 568

1951 AAAATGGTACCGCCGCACCGAGCCCTTCACTTCCCAGTGGCCTT..... 1994
       :::::::::   ||||||   :::   |||:::::|||
 568 isLeuAlaSerSerProProSerSerGlnAlaProGlyAlaLeuGlnGlu 584

1995 ......CCCACCAGGAGGAAAGCCATGTTCACAGCATATTATTTCTGTGA 2038
           ||||||        :::|||      |||
 585 CysProProThrLeuAlaProGlyMetThrLeuAla.............. 596

2039 CTGGATTTGTTGATAGTGACAGAGATGACCTAAAATTAATGGCTTATTTG 2088

 596 ..................................................  596

2089 GCAGGTGCCAAATATACGGGTTATCTATGCCGCAGCAACACAGTCCTCAT 2138
                                    |||   |||   :::   |||
 597 ............................ProValGlnGlyThrAlaHis 603

2139 CTGTAAAGAACCAACTGGTTTAAAGTATGAAAAAGCCAAAGAGTGGAGGA 2188
     :::   :::   :::                    :::         :::
 604 ValValLysGlyGly.....................AlaThrThrSerSerPr 614

2189 TACCCTGTGTCAACGCCCAGTGGCTTGGCGACATTCTTCTGGGAAACTTT 2238
     ::::::::::|||   |||:::|||:::      :::|||      |||
 614 oValValAlaGlnValProAlaAlaPheTyrMetGlnSerValHisLeuP 631

2239 GAGGC.........................ACTGAGGCAGATTCAGTATAG 2264
       |||                          :::||||||||||
 631 roGlyLysProGlnThrLeuAlaValLysArgLysAlaAspSerGluGlu 647

2265 TCGCTACACGGCATTCAGTCTGCAGGATCCATTTGCCCCTACCCAGCATT 2314
     :::          :::::::::   ||||||:::   |||      |||
 648 GluArgAspAspValSerThrLeuGlySerMetLeuProAlaLysAla.. 663

2315 TAGTTTTAAATCTTTTAGATGCTTGGAGAGTTCCCTTAAAAGTGTCTGCA 2364
                               |||   :::   :::      :
 664 ............................SerProValAlaGluSerProL 671

2365 GAGTTGTTGATGAGTATAAGACTACCTCCCAAACTGAAACAGAATGAAGT 2414
     ::|||:::||||||   |||::::::          ::::::||| :::||
 671 ysValMetAspGlu...LysSerSerLeuGlyGluLysAlaGlu.SerVa 686

2415 AGCTAATGTCCAG.........CCTTCTTCCCAAAGAGCCAGAATT.... 2451
```

```
          ||||||||||:::          ||||||||||:::   :::   :::
 686 lAlaAsnValAsnAlaAsnThrProSerSerGluLeuValAlaLeuThrP 703

2452 .....GAAGACGTACCACCTCCCACT..................... 2472
             :::|||||||||||||||||
 703 roAlaProSerValProProProThrLeuAlaMetValSerArgGlnMet 719

2473 .................................AAAAAGCTAACTCC 2486
                                      :::   ||||||
 720 GlyAspSerLysProProGlnAlaIleValLysProGlnIleLeuThrHi 736

2487 AGAATTGACCCCTTTTGTGCTTTTCACTGGATTCGAGCCTGTCCAGGTT. 2535
           :::      |||||||:::      |||   ||||||      |||
 736 sIleIleGluGlyPheValIleGlnGluGlyAlaGluProPheProValG 753

2536 .....CAACAGTATATTAAGAAG 2553
           :::|||   :::|||:::
 753 lyCysSerGlnLeuLeuLysGlu 760
```

seq_name: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-508-761B-22

```
seq_documentation_block:
; Sequence 22, Application US/08508761B
; Patent No. 6027920
;  GENERAL INFORMATION:
;    APPLICANT:  Joliff, Gwennael
;    APPLICANT:  Guyonvarch, Armel
;    APPLICANT:  Purification, Relano
;    APPLICANT:  Duchiron, Francis
;    APPLICANT:  Renaud, Michel
;    TITLE OF INVENTION:  System for Protein Expression and
;    TITLE OF INVENTION:  Secretion Especially in Corynebacteria
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Jacobson, Price, Holman & Stern, PLLC
;      STREET:  400 Seventh St. N.W.
;      CITY:  Washington D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/508,761B
;      FILING DATE:  31-JUL-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91/09652
;      FILING DATE:  29-JUL-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 91/09870
;      FILING DATE:  02-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Player, William E.
;      REGISTRATION NUMBER:  31,409
;      REFERENCE/DOCKET NUMBER:  P58525NA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 638-6666
;      TELEFAX:  (202) 393-5350
;  INFORMATION FOR SEQ ID NO:  22:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  120 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    ORIGINAL SOURCE:
;      ORGANISM:  synthetic
US-08-508-761B-22


alignment_scores:
```

```
           Quality:  173.00            Length:     97
             Ratio:  3.681              Gaps:      8
 Percent Similarity:  48.454   Percent Identity: 43.299

alignment_block:
 US-09-664-641-10/rev x US-08-508-761B-22   ..

 Align seg 1/1  to: US-08-508-761B-22  from: 1  to: 120

    1688 GCTGGATCATGTCCAAAAAGCTGATGCTGCTGAGGCGATGGT......GG 1645
         ||||||   |||      :::   ||||||   |||   |||
      21 AlaGlyCysCysThrGlyThrGlyCysCysThrGlyThrGlyCysCysTh 37

    1644 TGGCTGCTGTTGCTGC......TGCTGCTGCTGTGAAAATGGATGCGGCG 1601
         ||||||   ||||||      |||   ||||||   :::||||||||||
      37 rGlyCysGlyCysCysThrGlyCysGlyCysCys...ThrGlyCysGlyC 53

    1600 GCTGCTGGGGGTAAGGTTGC......TGAGAGATCTGCTGC......... 1566
           |||   |||   ||||||                ||||||
      53 ysCysThrGlyThrGlyCysCysThrGlyAlaGlyCysCysThrGlyGly 69

    1565 ...TGCTGC............TGCTGCTGGAGCTGCATTATCTGCTGGGG 1531
            ||||||            ||||||   :::|||      |||   ||
      70 GlyCysCysThrGlyGlyGlyCysCysThrGlyCysGlyCysCysThrGl 86

    1530 CTGAAGGTGT......AAAACCGGGTGCTGCTGC.........TGCTGCT 1496
         |      |||         ||||||||||||             ||||||
      86 yGlyGlyCysCysThrGlyThrGlyCysCysThrGlyGlyGlyCysCysT 103

    1495 GCTGGGCCTGCTGCTGCTGCTGTAGCATGTGTGTCTCTGGA 1455
           |||   ||||||      :::   |||   :::|||
     103 hrGlyAlaGlyCysCysThrGlyThrGlyCysCysThrGly 116

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-625-188-20

seq_documentation_block:
; Sequence 20, Application US/09625188
; Patent No. 6307037
; GENERAL INFORMATION:
;  APPLICANT:  No. 6307037artis AG
;  TITLE OF INVENTION:  Fungal Target Genes and Methods
;  FILE REFERENCE:  PB/5-31285P1
;  CURRENT APPLICATION NUMBER: US/09/625,188
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS:  44
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 20
;   LENGTH: 729
;   TYPE: PRT
;   ORGANISM: Ashbya gossypii
US-09-625-188-20


alignment_scores:
           Quality:  173.00            Length:    333
             Ratio:  1.036              Gaps:     15
 Percent Similarity:  50.150   Percent Identity: 25.225

alignment_block:
 US-09-664-641-10 x US-09-625-188-20   ..

 Align seg 1/1  to: US-09-625-188-20  from: 1  to: 729

     841 GAAGAAGAGGAAGAGGAAGAGGAGGAGGAAGTAGAAAATGAGGAACAAGA 890
         |||      :::|||      ::::::::::::      :::   |||
     227 GluValValAspGluLeuThrGlnGlnGlnIleProPheGlnThrGlnMe 243

     891 TTCTCAGAATGAGGGTAGTACAGATGAGAAGTCAAGCCCTGCCAGCTCTC 940
          ::::::::::::||||||          :::      ::::::   |
     243 tGlnMetHisGlnGlySerProProAlaProThrHisValThrAlaTyrG 260

     941 AAGAAGGGTCTCCTTCAGGTGACCAGCAGTTTTCACCTAAATCCAACACT 990
         ||:::        |||::::::        ||||||   ::::::::
     260 lnGlnProLeuLeuSerAlaSerGlyValValSerProProGlnSerAla 276

     991 GAAAAATCTAAA.....................GGGGAATTAATGTT 1016
         ::::::   |||                     :::::::::   ||
     277 SerSerValLysArgProSerLeuLeuValGlnArgSerArgPheSerPh 293

    1017 TGATGAT.........TCTTCAGATTCATCACCGGAAAAACAGGAGAGAA 1057
         |   |||         |||:::::::::::::      :::|||:::   :
     293 eThrAspProPheSerSerGluGlnThrAsnMetGlySerGlnGlnProA 310

    1058 ATTTAAACTGGACCCCGGCCGAAGTCCCACAGTTAGCTGCAGCAAAACGC 1107
         ::|||      ||||||   :::      |||   ::::::   :::
     310 spLeuIleThrThrProLeuLysGlyHisGlnAspThrGlyLysSerPhe 326

    1108 AGGCTGCCTCAGGGAAAGGAG........................... 1128
           |||   ::::::   |||
     327 LeuLeuMetGluSerAspGluIleAsnAspSerIleTrpGlyAsnGlyTh 343

    1129 .......CCTGGGTTGATTAAC...TTGTGTGCCAATGTCCCACCCGTCC 1168
             |||:::   |||:::   |||   ::::::   ||||||:::|
     343 rGlyIleProSerSerIleSerGlyLeuThrThrSerGlnProProThrP 360

    1169 CA.............................................. 1170
         ||
     360 roHisLeuGluTrpGlyThrThrGlyArgArgGlnSerSerThrPheTyr 376

    1171 ............GGTAACATTTTGCCCCCT..............GAGGT 1193
                    ::::::   :::||||||               :::||
     377 ProSerGlnSerAsnThrGluIleProProMetHisLeuThrGlyGlnVa 393

    1194 CCGGGGTAATTTAATGGCTGCTGGA......CAAAACCTCCAAAGTTCTG 1237
         |:::::::::   :::|||:::|||      |||   |||   :::   :
     393 lGlnSerSerGlnLeuAlaThrGlyLeuGlnGlnProLeuProGlnProG 410

    1238 AAAGATCAGAAATGATAGCTACCTGGAGTCCAGCTGTACGGACACTGAGG 1287
         ::|||:::::::::                  :::   :::   :::
     410 lnArgGlnSerLeu.................SerTyrAsnLeuValThr 420

    1288 AATATTACTAATAATGCTGACATTCAGCAGATGAACCGGCCA......TC 1331
         ::::::::::::                |||||   |||      ||
     421 ProLeuSerSerAsp..............MetAsnLeuProProGlnSe 432

    1332 AAATGTAGCACATATCTTA...CAGACTCTTTCAGCACCTACGAAAAATT 1378
         |:::   :::   ||||||   :::          |||   |||:::
     432 rSerGlnGlyGlyIleLeuProHisGlnAlaProAlaGlnThrGlnProG 449

    1379 TAGAACAGCAGGTGAATCACAGCCAGCAGGGACATACAAATGCCAATGCA 1428
           :::|||   ::::::|||   |||:::   |||
     449 lnSerGlnAlaLeuGlnHisHisGlnHisLeuHis.............. 460

    1429 GTGCTGTTTAGCCAAGTGAAAGTGACTCCAGAGACACACATGCTACAGCA 1478
               :::|||   ::::::      :::   |||   ||||||||
     461 ......HisGlnGlnGlnGlnLeuGlnGlnGlnGlnHisHisLeuGlnGl 475

    1479 GCAGCAGCAGGCCCAGCAGCAGCAGCAGCAGCACCCGGTTTTACACCTTC 1528
         |||||||:::   |||||||||||||:::                |||      |
     475 nGlnGlnHisGlnGlnGlnGlnGlnSer.............LeuSerGlnG 488

    1529 AGCCCCAGCAGATAATGCAGCTCCAG..............CAGCAGCAG 1563
         ||||||||||   ::::::   |||                ||||||:::
     488 lnProGlnGlnGlnGlnSerGlnGlnSerGlnAlaHisSerGlnGlnHis 504

    1564 CAGCAGCAGATCTCTCAGCAACCTTACCCCCAGCAGCCGCCGCATCCATT 1613
         |||||||||   :::||||||          |||||||||
     505 GlnGlnGlnHisGlnGlnGlnGlnGlnGlnGlnGlnPro............ 517

    1614 TTCACAGCAGCAGCAGCAGCAACAGCAGCCACCACCATCGCCTCAGCAG 1662
         ||||||||||||   ||||||:::||||||   :::|||||||||
     518 ....GlnGlnGlnGlnProGlnGlnHisProProGlnGlnProGlnGln 532
```

OM of: US-09-664-641-10 to: A_Geneseq_032802:* out_format : pfs

Date: Jun 4, 2002 5:32 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09664641/runat_04062002_110932_9158/app_query.fasta_1.3697
-DB=A_Geneseq_032802 -QFMT=fastan -SUFFIX=rag -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09664641_@CGN1_1_206 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-664-641-10
Query length: 3580
Database: A_Geneseq_032802:*
Database sequences: 747574
Database length: 111073796
Search time (sec): 168.130000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..			
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU27822	+	4791.00	8230.72	0.0	91				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB71695	+	852.00	1419.51	3.5e-72	17				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU27994	+	319.00	535.15	1.1e-21	10				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAO05204	+	294.50	488.80	3.1e-19	14				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY03183	+	263.50	405.29	1.3e-15	15				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG23599	-	249.00	421.02	4.3e-15	61				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG19008	-	248.00	395.52	1.7e-14	39				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG23600	-	242.00	404.43	2.5e-14	87				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU07343	-	238.00	360.00	3.9e-13	16				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB69280	+	225.50	348.20	3.8e-12	77				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB63516	+	220.50	334.19	1.5e-11	11				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG27446	-	220.00	366.15	3.3e-12	89				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY60558	-	216.00	351.12	1.2e-11	16				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB61245	+	216.00	336.94	2.4e-11	51				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG27784	-	215.50	358.24	9.1e-12	90				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY96255	+	214.00	323.22	6.3e-11	11				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY58500	+	214.00	323.22	6.3e-11	11				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB62331	+	214.00	323.22	6.3e-11	11				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB63057	+	213.50	313.79	1.1e-10	22				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB68896	+	208.50	304.32	3.4e-10	24				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG27074	-	206.50	338.84	8.1e-11	12				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG23598	-	206.50	338.03	8.4e-11	13				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG27785	-	206.50	338.03	8.4e-11	13				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB67633	+	206.50	313.16	2.9e-10	92				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB69615	+	205.00	310.45	4.0e-10	93				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY54320	+	200.00	292.02	2.0e-09	20				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB67519	+	199.50	303.55	1.2e-09	76				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAM41453	+	199.00	311.22	8.8e-10	38				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY54319	+	199.00	289.97	2.5e-09	20				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB58382	+	198.50	282.47	3.9e-09	35				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG17148	+	198.00	284.63	3.7e-09	27				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB72020	+	197.50	302.47	1.7e-09	63				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB70437	+	197.00	280.73	5.2e-09	32				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG27789	-	196.50	324.61	6.4e-10	96				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB84634	+	195.50	278.47	7.1e-09	31				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG29282	-	194.50	325.34	8.1e-10	69				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG27787	-	194.50	312.77	1.5e-09	18				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB62018	+	194.50	292.78	4.0e-09	90				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB69365	+	193.50	288.23	5.8e-09	11				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR38746	+	193.00	296.62	4.1e-09	54				
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAM39667	+	193.00	295.85	4.3e-09					
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG27786	-	192.00	318.08	1.6e-09					
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB30808	+	192.00	303.00	3.4e-09					
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB61004	+	192.00	289.00	6.8e-09					
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB64877	+	192.00	272.87	1.5e-08					

seq_name: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU27822

seq_documentation_block:
1D AAU27822 standard; Protein; 911 AA.
XX
AC AAU27822;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human full-length polypeptide sequence #147.
XX
KW Mammal; human; rhesus monkey; baker's yeast; fission yeast; Norway rat;
KW mouse; Chinese hamster; African clawed frog; fruit fly; dog; leukaemia;
KW cancer; lymphoma; neuroblastoma; autoimmune disorder; cell proliferation;
KW nervous system disorder; inflammatory disorder; cell differentiation;
KW angiogenesis; stem cell growth factor; activin; inhibin; cartilage; burn;
KW genetic disorder; bone regeneration; tendon; ligament; tissue repair;
KW cytostatic; antirheumatic; antiarthritic; vulnerary; antiinflammatory;
KW antibacterial; immunosuppressive; vasotropic; antiparkinsonian;
KW neuroprotective; osteopathic; antidiabetic; antiasthmatic; antiallergic;
KW immunostimulant; analgesic; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200164834-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US04926.
XX
PR 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
PR 17-JUN-2000; 2000US-0597707.
PR 14-JUL-2000; 2000US-0616807.
PR 19-SEP-2000; 2000US-0664641.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Wang J, Ma Y, Wang D, Chen R, Xu C;
PI Drmanac R;
XX
DR WPI; 2001-589862/66.
DR N-PSDB; AAS44722.
XX
PT Novel polypeptides and nucleic acids obtained from cDNA libraries
PT prepared from various human tissues, for diagnosis, treatment of
PT cancer, neurological, inflammatory disorders and for use in arrays for
PT detection -
XX
PS Claim 10; SEQ ID No 319; 153pp; English.
XX
CC Sequences AAU27676-AAU28019 represent full-length polypeptides and
CC contig polypeptides of the invention. The proteins and their associated
CC DNA sequences are useful for the treatment, diagnosis and prevention of
CC various types of disorder in a mammalian subject such as a human, dog,
CC monkey, mouse, hamster or rat. The disorders include cancers such as
CC leukaemia, lymphoma and neuroblastoma, autoimmune disorders such as
CC multiple sclerosis, connective tissue disease, rheumatoid arthritis,
CC diabetes mellitus, allergic rhinitis, asthma and eczema, nervous system
CC disorders such as Parkinson's disease, Alzheimer's disease, Huntington's
CC chorea, amyotrophic lateral sclerosis, spinal muscular atrophy and
CC Wernicke disease, inflammatory disorders such as nephritis, Crohn's
CC disease, ischaemia-reperfusion injury, shock, sepsis and inflammatory
CC bowel disease. The sequences exhibit activity relating to angiogenesis,
CC cell proliferation, cell differentiation, stem cell growth factor,
CC activin or inhibin. Therefore, they can be used to manipulate stem cells

```
CC   in culture to give rise to neuroepithelial cells that can be used to
CC   augment or replace cells damaged by illness, accidental damage or genetic
CC   disorders. The sequences may also be used for regeneration of bone,
CC   cartilage, tendons and ligaments and in tissue repair and burn healing.
CC   Note: Some sequences for this patent did not form part of the printed
CC   specification, but were obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   911 AA;


alignment_scores:
            Quality: 4791.00              Length:    911
              Ratio:  5.259                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-664-641-10 x AAU27822   ..

 Align seg 1/1  to: AAU27822  from: 1  to: 911

     373 ATGGTTTTTCTCCAGAATCATGTCAGATTTTTTTTGGAATCACTGCCTGC 422
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetValPheLeuGlnAsnHisValArgPhePheLeuGluSerLeuProAl 17

     423 CTTTCTCAGGGTGTTGATACAAGCTGGAGCTCTTTGTTGGAGTCTTCCAG 472
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      17 aPheLeuArgValLeuIleGlnAlaGlyAlaLeuCysTrpSerLeuProG 34

     473 AGCTCTCCCAAGGAGAGGTAGGGAAGGGAGCTTGTCCAGCAGAAGTTGGG 522
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 luLeuSerGlnGlyGluValGlyLysGlyAlaCysProAlaGluValGly 50

     523 AAGCACAGAGATCATCTGCCTTCTTCTGACCCGGTATTGATGCAGGCTGA 572
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 LysHisArgAspHisLeuProSerSerAspProValLeuMetGlnAlaGl 67

     573 GGCCTCTGTTGTAATGTGCTGGGTGTCATCTGAAGACAGAAGTGCCCTGT 622
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 uAlaSerValValMetCysTrpValSerSerGluAspArgSerAlaLeuT 84

     623 GGGCTTTGGTTACGTTCTATGGGGGAGATTGCCAGCTAACCCTCAATAAG 672
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      84 rpAlaLeuValThrPheTyrGlyGlyAspCysGlnLeuThrLeuAsnLys 100

     673 AAATGCACGCATTTGATTGTTCCAGACCAAAGGGGGGGAAATACGAATG 722
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 LysCysThrHisLeuIleValProGluProLysGlyGlyLysTyrGluCy 117

     723 TGCTTTAAAGCGAGCAAGTATTAAAATTGTGACTCCTGACTGGGTTCTGG 772
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     117 sAlaLeuLysArgAlaSerIleLysIleValThrProAspTrpValLeuA 134

     773 ATTGCGTATCAGAGAAAACCAAAAAGGACGAAGCATTTTATCATCCTCGT 822
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     134 spCysValSerGluLysThrLysLysAspGluAlaPheTyrHisProArg 150

     823 CTGATTATTTATGAAGAGGAAGAAGAGGAAGAGGAAGAGGAGGAGGAAGT 872
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 LeuIleIleTyrGluGluGluGluGluGluGluGluGluGluGluGluVa 167

     873 AGAAAATGAGGAACAAGATTCTCAGAATGAGGGTAGTACAGATGAGAAGT 922
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     167 lGluAsnGluGluGlnAspSerGlnAsnGluGlySerThrAspGluLysS 184

     923 CAAGCCCTGCCAGCTCTCAAGAAGGGTCTCCTTCAGGTGACCAGCAGTTT 972
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     184 erSerProAlaSerSerGlnGluGlySerProSerGlyAspGlnGlnPhe 200

     973 TCACCTAAATCCAACACTGAAAAATCTAAAGGGGAATTAATGTTTGATGA 1022
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 SerProLysSerAsnThrGluLysSerLysGlyGluLeuMetPheAspAs 217

    1023 TTCTTCAGATTCATCACCGGAAAAACAGGAGAGAAATTTAAACTGGACCC 1072
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     217 pSerSerAspSerSerProGluLysGlnGluArgAsnLeuAsnTrpThrP 234

    1073 CGGCCGAAGTCCCACAGTTAGCTGCAGCAAAACGCAGGCTGCCTCAGGGA 1122
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     234 roAlaGluValProGlnLeuAlaAlaAlaLysArgArgLeuProGlnGly 250

    1123 AAGGAGCCTGGGTTGATTAACTTGTGTGCCAATGTCCCACCCGTCCCAGG 1172
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 LysGluProGlyLeuIleAsnLeuCysAlaAsnValProProValProGl 267

    1173 TAACATTTTGCCCCCTGAGGTCCGGGGTAATTTAATGGCTGCTGGACAAA 1222
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     267 yAsnIleLeuProProGluValArgGlyAsnLeuMetAlaAlaGlyGlnA 284

    1223 ACCTCCAAAGTTCTGAAAGATCAGAAATGATAGCTACCTGGAGTCCAGCT 1272
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     284 snLeuGlnSerSerGluArgSerGluMetIleAlaThrTrpSerProAla 300

    1273 GTACGGACACTGAGGAATATTACTAATAATGCTGACATTCAGCAGATGAA 1322
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     301 ValArgThrLeuArgAsnIleThrAsnAsnAlaAspIleGlnGlnMetAs 317

    1323 CCGGCCATCAAATGTAGCACATATCTTACAGACTCTTTCAGCACCTACGA 1372
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     317 nArgProSerAsnValAlaHisIleLeuGlnThrLeuSerAlaProThrL 334

    1373 AAAATTTAGAACAGCAGGTGAATCACAGCCAGCAGGGACATACAAATGCC 1422
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     334 ysAsnLeuGluGlnGlnValAsnHisSerGlnGlnGlyHisThrAsnAla 350

    1423 AATGCAGTGCTGTTTAGCCAAGTGAAAGTGACTCCAGAGACACACATGCT 1472
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     351 AsnAlaValLeuPheSerGlnValLysValThrProGluThrHisMetLe 367

    1473 ACAGCAGCAGCAGCAGGCCCAGCAGCAGCAGCAGCAGCACCCGGTTTTAC 1522
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     367 uGlnGlnGlnGlnGlnAlaGlnGlnGlnGlnGlnGlnHisProValLeuH 384

    1523 ACCTTCAGCCCCAGCAGATAATGCAGCTCCAGCAGCAGCAGCAGCAGCAG 1572
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     384 isLeuGlnProGlnGlnIleMetGlnLeuGlnGlnGlnGlnGlnGlnGln 400

    1573 ATCTCTCAGCAACCTTACCCCCAGCAGCCGCCGCATCCATTTTCACAGCA 1622
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     401 IleSerGlnGlnProTyrProGlnGlnProProHisProPheSerGlnGl 417

    1623 GCAGCAGCAGCAACAGCAGCCACCACCATCGCCTCAGCAGCATCAGCTTT 1672
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     417 nGlnGlnGlnGlnGlnProProProSerProGlnGlnHisGlnLeuP 434

    1673 TTGGACATGATCCAGCAGTGGAGATTCCAGAAGAAGGCTTCTTATTGGGA 1722
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     434 heGlyHisAspProAlaValGluIleProGluGluGlyPheLeuLeuGly 450

    1723 TGTGTGTTTGCAATTGCGGATTATCCAGAGCAGATGTCTGATAAGCAACT 1772
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     451 CysValPheAlaIleAlaAspTyrProGluGlnMetSerAspLysGlnLe 467

    1773 GCTGGCCACCTGGAAAAGGATAATCCAGGCACATGGCGGCACTGTTGACC 1822
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     467 uLeuAlaThrTrpLysArgIleIleGlnAlaHisGlyGlyThrValAspP 484

    1823 CCACCTTCACGAGTCGATGCACGCACCTTCTCTGTGAGAGTCAAGTCAGC 1872
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     484 roThrPheThrSerArgCysThrHisLeuLeuCysGluSerGlnValSer 500

    1873 AGCGCGTATGCACAGGCAATAAGAGAAAGAAAGAGATGTGTTACTGCACA 1922
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     501 SerAlaTyrAlaGlnAlaIleArgGluArgLysArgCysValThrAlaHi 517

    1923 CTGGTTAAACACAGTCTTAAAAAAAAAAAAAAATGGTACCGCCGCACCGAG 1972
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  517 sTrpLeuAsnThrValLeuLysLysLysLysMetValProProHisArgA 534

 1973 CCCTTCACTTCCCAGTGGCCTTCCCACCAGGAGCGAAAGCCATGTTCACAG 2022
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  534 laLeuHisPheProValAlaPheProProGlyGlyLysProCysSerGln 550

 2023 CATATTATTTCTGTGACTGGATTTGTTGATAGTGACAGAGATGACCTAAA 2072
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  551 HisIleIleSerValThrGlyPheValAspSerAspArgAspAspLeuLy 567

 2073 ATTAATGGCTTATTTGGCAGGTGCCAAATATACGGGTTATCTATGCCGCA 2122
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  567 sLeuMetAlaTyrLeuAlaGlyAlaLysTyrThrGlyTyrLeuCysArgS 584

 2123 GCAACACAGTCCTCATCTGTAAAGAACCAACTGGTTTAAAGTATGAAAAA 2172
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  584 erAsnThrValLeuIleCysLysGluProThrGlyLeuLysTyrGluLys 600

 2173 GCCAAAGAGTGGAGGATACCCTGTGTCAACGCCCAGTGGCTTGGCGACAT 2222
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  601 AlaLysGluTrpArgIleProCysValAsnAlaGlnTrpLeuGlyAspIl 617

 2223 TCTTCTGGGAAACTTTGAGGCACTGAGGCAGATTCAGTATAGTCGCTACA 2272
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  617 eLeuLeuGlyAsnPheGluAlaLeuArgGlnIleGlnTyrSerArgTyrT 634

 2273 CGGCATTCAGTCTGCAGGATCCATTTCCCCCTACCCAGCATTTAGTTTTA 2322
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  634 hrAlaPheSerLeuGlnAspProPheAlaProThrGlnHisLeuValLeu 650

 2323 AATCTTTTAGATGCTTGGAGAGTTCCCTTAAAAGTGTCTGCAGAGTTGTT 2372
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  651 AsnLeuLeuAspAlaTrpArgValProLeuLysValSerAlaGluLeuLe 667

 2373 GATGAGTATAAGACTACCTCCCAAACTGAAACAGAATGAAGTAGCTAATG 2422
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  667 uMetSerIleArgLeuProProLysLeuLysGlnAsnGluValAlaAsnV 684

 2423 TCCAGCCTTCTTCCCAAAGAGCCAGAATTCAA.ACGTACCACCTCCCACT 2472
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  684 alGlnProSerSerGlnArgAlaArgIleGluAspValProProProThr 700

 2473 AAAAAGCTAACTCCAGAATTGACCCCTTTTGTGCTTTTCACTGGATTCGA 2522
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  701 LysLysLeuThrProGluLeuThrProPheValLeuPheThrGlyPheGl 717

 2523 GCCTGTCCAGGTTCAACAGTATATTAAGAAGCTCTACATTCTTGGTGGAG 2572
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  717 uProValGlnValGlnGlnTyrIleLysLysLeuTyrIleLeuGlyGlyG 734

 2573 AGGTTGCGGAGTCTGCACAGAAGTGCACACACCTCATTGCCAGCAAAGTG 2622
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  734 luValAlaGluSerAlaGlnLysCysThrHisLeuIleAlaSe LysVal 750

 2623 ACTCGCACCGTGAAGTTCCTGACGGCGATTTCTGTCGTGAAGCACATAGT 2672
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  751 ThrArgThrValLysPheLeuThrAlaIleSerValValLysHisIleVa 767

 2673 GACGCCAGAGTGGCTGGAAGAATGCTTCAGGTGTCAGAAGTTCATTGATG 2722
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  767 lThrProGluTrpLeuGluGluCysPheArgCysGlnLysPheIleAspG 784

 2723 AGCAGAACTACATTCTCCGACATGCTGAGGCAGAAGTACTTTTCTCTTTC 2772
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  784 luGlnAsnTyrIleLeuArgAspAlaGluAlaGluValLeuPheSerPhe 800

 2773 AGCTTGGAAGAATCCTTAAAACGGGCACACGTTTCTCCACTCTTTAAGGC 2822
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  801 SerLeuGluGluSerLeuLysArgAlaHisValSerProLeuPheLysAl 817

 2823 AAAATATTTTTACATCACACCTGGAATCTGCCCAAGTCTTTCCACTATGA 2872
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
  817 aLysTyrP: TyrIleThrProGlyIleCysProSerLeuSerThrMetL 834

 2873 AGGCAATCGTAGAGTGTGCAGGAGGAAAGGTGTTATCCAAGCAGCCATCT 2922
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  834 ysAlaIleValGluCysAlaGlyGlyLysValLeuSerLysGlnProSer 850

 2923 TTCCGGAAGCTCATGGAGCACAAGCAGAACTCGAGTTTGTCGGAAATAAT 2972
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  851 PheArgLysLeuMetGluHisLysGlnAsnSerSerLeuSerGluIleL 867

 2973 TTTAATATCCTGTGAAAATGACCTTCATTTATGCCGAGAATATTTTGCCA 3022
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  867 eLeuIleSerCysGluAsnAspLeuHisLeuCysArgGluTyrPheAlaA 884

 3023 GAGGCATAGATGTTCACAATGCAGAGTTCGTTCTGACTGGAGTGCTCACT 3072
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  884 rgGlyIleAspValHisAsnAlaGluPheValLeuThrGlyValLeuThr 900

 3073 CAAACGCTGGACTATGAATCATATAAGTTTAAC 3105
                 ||||||||||||||||||||||||||||||||
  901 GlnThrLeuAspTyrGluSerTyrLysPheAsn 911
```

seq_name: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB71695

seq_documentation_block:

```
ID   ABB71695 standard; Protein; 1798 AA.
XX
AC   ABB71695;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 41877.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL15798.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 41877; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1798 AA;
```

```
alignment_scores:
           Quality:  852.00              Length:    848
             Ratio:  1.906                Gaps:     24
 Percent Similarity:  52.712   Percent Identity: 28.774

alignment_block:
 US-09-664-641-10 x ABB71695   ..

 Align seg 1/1  to: ABB71695  from: 1  to: 1798

     925 AGCCCTGCCAGCTCTCAAGAAGGGTCTCCTTCAGGTGACCAGCAGTTTTC 974
         ||||||   :::::::|||   |||:::         :::||||||   ||
     974 SerProIleGlnGlnGlnProGlyAlaGlyProGlnGlnGlnGlnGlnSe 990

     975 ACCTAAATCCAACACTGAAAAATCTAAAGGGCAATTAATGTTTGATGATT 1024
         ||||::::::::::      :::  :::                |||  :
     990 rProGlnGlnGlnProPheAsnProAsn.............AspProA 1002

    1025 CTTCAGATTCATCACCGGAAAAACAGGAGACAAATTTAAACTGGACCCCG 1074
         ::::::::        ::::::|||:::  :::::::::      |||
    1002 snGlnGlnMetLeuMetGlnArgGlnGlnLeuArgValGlnGlnIlePro 1018

    1075 GCCGAA.........GTCCCACAGTTAGCTGCAGCAAAACGCAGGCTGCC 1115
            :::             ||||||:::           :::|||
    1019 IleGlnGlnGlnGlyProGlnGlnLeuGlyLysGlnThrValGlnLeuIl 1035

    1116 TCAGGGAAAGGAGCCTGGGTTGATT........................... 1140
         |||  :::     ||||||
    1035 eThrGlyProAsnGlyGlnLeuIleGluGlnGlnThrIleValGlnGlnP 1052

    1141 ........AACTTGTGTGCCAATGTCCCACCCGTCCCA......GGTAAC 1176
                 :::  :::  :::::::|||  :::|||     |||:::
    1052 roGlnGlnGlnSerAlaProGlyThrProThrThrProGlyAlaGlyGly 1068

    1177 ATTTTGCCCCCTGAGGTCCGGGGTAATTTAATGGCTGCTGGACAAAACCT 1226
         :::|||||||           ||||||                |||:::||
    1069 ValLeuPro............GlyAsnAlaAsnIleMetThrGlnThrLe 1081

    1227 C...CAAAGTTCTGAAAGATCAGAAATGATAGCTACCTGGAGTCCAGCTG 1273
         |   :::::::|||          :::                ||||||
    1081 uValMetThrSerThrThrProAspGlyGlnProGlnGlnSerProGlnG 1098

    1274 TACGGACACTGAGGAATATTACTAATAATGCTGACATTCAGCAGATGAAC 1323
            :::|||::::::   |||      |||:::::   :::   :::
    1098 lnMetAsnLeuLysThrLysThrAlaLeuAlaAsnMetLeuSerAsnArg 1114

    1324 CGGCCATCAAATGTAGCACATATCTTACAGACTCTTTCAGCACCTACGAA 1373
            :::|||   |||:::      |||
    1115 LeuGlyAsnAsnGlyAlaGlnThrProGln..................... 1124

    1374 AAATTTAGAACAGCAGGTGAATCACAGCCAGCAGGGACATACAAATGCCA 1423
                 ||||||:::   :::      :::|||:::      :
    1125 .........GlnGlnIleIleGlnLeuProAspGlyGlnGlnProGlnG 1138

    1424 ATGCAGTGCTGTTTAGCCAAGTGAAAGTGACT.............CCA 1458
         ::      :::   :::|||||   ||||||
    1138 lnGlnGlnIleValGlnGlnValValValThrGlyAlaIleGlnGlnGln 1154

    1459 GAGACACACATGCTACAGCAGCAGCAGCAGGCCCAGCAGCAGCAGCAGCA 1508
         :::   ::::::|||   ||||||||||||   ||||||||||||||||
    1155 GlnGlnGlnLeuLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGl 1171

    1509 G...................CACCCGGTTTTACACCTTCAGCCC.... 1533
         |                    |||      |||:::::::|||
    1171 nMetValValGlnGlnGlnLeuProProGlnGlnHisValGluProProS 1188

    1533 ..................................................... 1533

    1188 erAlaAlaGlyAlaLeuArgMetMetGlyGlnGlnHisAsnAlaThrAla 1204

    1534 .......................CAGCAGATAATGCAGCTCCAGCA 1556

                                    |||:::::::::::::||||||||
    1205 GlyValProGlyProProArgThrGlnGlnGluLeuLeuMetLeuGlnGl 1221

    1557 GCAGCAGCAGCAGCAGATCTCTCAGCAA..................... 1584
         ||||||||||||||||:::   ||||||
    1221 nGlnGlnGlnGlnGlnLeuLeuGlnGlnGlnGlnGlnMetArgArgValValA 1238

    1585 .....CCTTACCCCCAGCAGCCGCCGCATCCATTT.............. 1614
              |||      ||||||   |||:::
    1238 laLeuProGlnGlnGlnGlnAlaProGlnGlnHisLeuValGlnGlnGln 1254

    1615 ....................TCACAGCAGCAGCAGCAGCAA........ 1635
                             |||   ||||||||||||||
    1255 IleValValAlaProThrGlnSerProGlnGlnGlnGlnGlnMetAlaVa 1271

    1635 ..................................................... 1635

    1271 lGlyValGlyValGlyValProValGlnArgThrProHisGlyTyrIleG 1288

    1636 ..............CAGCAGCCACCACCATCGCCTCAGCAGCATCAGCTT 1671
                      |||||||||   |||   |||      |||:::
    1288 lnThrArgProGlyGlnGlnProIleProMetPro.........GlnPhe 1301

    1672 TTTGGACATGATCCAGCAGTGGAGATTCCAGAAGAAGGCTTCTTATTGGG 1721
         :::|||||:::|||   ::::::::::|||   :::   ||||||:::||
    1302 TyrGlyHisAsnProAsnLeuLysLeuProAlaAspLeuPheLeuValGl 1318

    1722 ATGTGTGTTTGCAATTGCGGATTATCCAGAGCAGATGTCTGATAAGCAAC 1771
         ||||:::|||   |||::::::::|||   |||      :::|||:::
    1318 yCysThrPheTyrIleValGluTyrAspGluThrAspGlyAspGlu.... 1333

    1772 TGCTGGCCACCTGGAAAAGGATAATCCAGGCACATGGCGGCACTGTTGAC 1821
           |||      |||         |||:::      ||||||   ::::::
    1334 ..LeuProIleTrpLeuAspThrIleArgGlnPheGlyGlyAspIleGlu 1349

    1822 CCCACCTTCACGAGTCGATGCACGCACCTTCTCTGTGAGAGTCAAGTCAG 1871
            ::::::      |||  |||||||:::::|||:::::|||
    1350 ArgValTyrC sProArgValThrHisValIleCysArgThrGlnArgHi 1366

    1872 CAGCGCGTATGCACAGGCAATAAGAGAAAGAAAGAAGATGTGTTACTGCAC 1921
            ::::::      ||||||:::|||:::   |||||||||||||||||:
    1366 sGlyValValMetGlnAlaLeuArgAspAlaLysArgCysValThrAlaT 1383

    1922 ACTGGTTAAACACAGTCTTAAAAAAAAAAAAAAAATGGTACCGCCGCACCGA 1971
         ::|||||:::   :::      |||:::::::::::::|||||   :::
    1383 yrTrpLeuSerAspIleCysLeuLysArgGlnLeuMetProProTrpGln 1399

    1972 GCCCTTCACTTCCCAGTGGCCTTCCCACCAGGA.........GGAAAGCC 2012
            ||||||:::|||      ||||||                ||||||
    1400 ProLeuHisLeuPro......PheProSerGlnPheGlyTyrArgLysPr 1414

    2013 ATGTTCACAGCATATTATTTCTGTGACTGGATTTGTTGATAGTGACAGAG 2062
         |   ::::::::|||||||:::      ||||||      ::::::
    1414 oLeuGluArgTyrIleIleThrSerGluGlyPheGluGlyGluGluValV 1431

    2063 ATGACCTAAAATTAATGGCTTATTTGGCAGGTGCCAAATATACGGGTTAT 2112
            |||:::  ||||||      :::||||||  ||||||:::|||
    1431 alArgLeuGlnGlnMetAlaGluGluCysGlyAlaIleTyrThrSerTyr 1447

    2113 CTATGCCGCAGCAACACAGTCCTCATCTGTAAAGAACCAACTGGTTTAAA 2162
         |||   :::   ||||||||:::::|||||:::      |||   ||
    1448 LeuSerI /sValAsnThrValValValCysLysGlnLeuGluGlyAsnLy 1464

    2163 GTATGAAAAAGCCAAAGAGTGGAGGATACCCTGTGTCAACGCCCAGTGGC 2212
         |:::::   ||||||||||||:::||||||   |||||||||   ||||
    1464 sPheAsnAlaAlaLysGluTrpAsnIleProMetValAsnAlaLeuTrpL 1481

    2213 TTGGCGACATTCTTCTGGGAAACTTTGAGGCACTGAGGCAGATTCAGTAT 2262
         ||:::|||:::   :::||||||:::::::::|||   |||   :::
    1481 euSerAspValCysIleGlyAsnLeuSerGlyLeuSerGlnTyrGluAsn 1497

    2263 AGTCGCTACACGGCATTCAGTCTGCAGGATCCATTTGCCCCTACCCAGCA 2312
            :::|||      ::::::|||      ||||||          ::
```

```
 1498 ProLysTyrGlnGlnTyrAsnLeuValAlaProPheArgIleGluCysAs 1514

 2313 TTTAGTTTTAAATCTTTTAGATGCTTGGAGAGTTCCCTTAAAAGTGTCTG 2362
      :||||||   :::||||||   ||||||:::::::|||:::::::::::::
 1514 nLeuValAlaHisLeuLeuThrAlaTrpLysAlaProIleAsnLeuThrG 1531

 2363 CAGAGTTGTTGATGAGTATAAGACTA..................... 2388
        |||           ::::::
 1531 lnGluAlaHisGluArgValLysArgHisLeuSerAspProTyrGlyAsn 1547

 2389 CCTCCCAAACTGAAACAGAATGAAGTAGCTAATGTCCAGCCTTCTTCCCA 2438
      ||||||||||:::::::::::::        |||   ::::::
 1548 GluGlnLysLeuLysArgGlnLysMetAsnAlaPheGlnGlnGluGlnLe 1564

 2439 AAGAGCCAGAATTGAAGACGTACCACCTCCCACTAAAAAGCTAACTCCAG 2488
      :::|||     |||     ||||||           |||:
 1564 uProGluGlnIleValCysValGluTyrProThrThrThrLysProProL 1581

 2489 AATTGACCCCTTTTGTGCTTTTCACTGGATTCGAGCCTGTCCAGGTTCAA 2538
      :::::                    |||:::  |||
 1581 ysValIle.....................PheSerGlnValAlaAspAla 1590

 2539 CAGTATATTAAGAAG...CTCTACATTCTTGGTGGAGAGGTTGCGGAGTC 2585
      :::   :::|||||   :::   ||||||||||||   |||:::::::||
 1591 GluAlaLeuLysLysAlaValLeuIleLeuGlyGlyIleValValAspSe 1607

 2586 TGCACAGAAGTGCACACACCTCATTGCCAGCAAAGTGACTCGCACCGTGA 2635
      |          :::||||||||||:::   ::::::   :::||||||   |
 1607 rProAlaAspAlaThrHisLeuValMetThrArgGluSerArgThrCysL 1624

 2636 AGTTCCTGACGGCGATTTCTGTCGTGAAGCACATAGTGACGCCAGAGTGG 2685
      ||::::::   |||         |||   :::::::::      :::|||
 1624 ysLeuIleGlnAlaCysCysHisValAspTyrValLeuLysSerSerTrp 1640

 2686 CTGGAAGAATGCTTCAGGTGTCAGAAGTTCATTGATGAGCAGAACTACAT 2735
      :::   :::      ::::::   |||||||:::     :::   |||
 1641 IleAlaAspSerAlaLysAlaGlyLysPheValProThrAspProTyrAr 1657

 2736 TCTCCGAGATGCTGAGGCAGAAGTACTTTTCTCTTTCAGCTTGGAAGAAT 2785
      ::::::         ::::::      ::::::|||:::|||:::
 1657 qIleGlnHisIleProValAspGluAsnLeuGlnPheAsnLeuAsnThrV 1674

 2786 CCTTAAAACGGGCACACGTTTCTCCACTCTTTAAGGCAAAATATTTTTAC 2835
        |||               |||   ||||||   :::||||||||||:::
 1674 alLeuCysAlaProThrArgSerThrLeuPheAlaGlyLysTyrPheHis 1690

 2836 ATCACACCTGGAATCTGCCCAAGTCTTTCCACTATGAAGGCAATCGTAGA 2885
      :::||||||   :::   |||:::   :::   :::      :::::::||
 1691 ValThrProAspValPheProAlaArgGluGluIleIleArgMetIleGl 1707

 2886 GTGTGCAGGAGGAAAGGTGTTATCCAAGCAGCCATCTTTCCGGAAGCTCA 2935
      |   :::||||||||||||      |||:::   |||      ::::::
 1707 uSerSerGlyGlyLysValGluProLysArgArgSerGlyAlaSerValA 1724

 2936 TGGAG...CACAAGCAGAACTCGAGTTTGTCGGAAATAATTTTAATATCC 2982
        |||   |||   |||      :::   ::::::   |||:::::::::
 1724 laGluThrHisMetGln......AlaProAspSerTyrIleIleValThr 1738

 2983 TGTGAAAATGACCTTCATTTATGCCGAGAATATTTTGCCAGAGGC..... 3027
      |||   :::|||:::||||||||||   :::         :::|||
 1739 CysProThrAspMetHisLeuCysAlaAspLeuThrArgHisGlyAsnPr 1755

 3028 ....ATAGATGTTCACAATGCAGAGTTCGTTCTGACTGGAGTGCTCACTC 3073
              :::   ::::::||||||||:::::::::::|||   |
 1755 oLysCysHisIleValSerThrGluPheValMetSerSerIleLeuArgG 1772

 3074 AAACGCTGGACTATGAA............TCATATAAGTTTAAC 3105
      ||   |||:::   |||           |||   :::|||
 1772 lnGlnLeuGluIleGluProAsnLeuIleProTyrLeuTyrAsn 1786

seq_name: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU27994

seq_documentation_block:
```

```
ID   AAU27994 standard; Protein; 104 AA.
XX
AC   AAU27994;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Human contig polypeptide sequence #147.
XX
KW   Mammal; human; rhesus monkey; baker's yeast; fission yeast; Norway rat;
KW   mouse; Chinese hamster; African clawed frog; fruit fly; dog; leukaemia;
KW   cancer; lymphoma; neuroblastoma; autoimmune disorder; cell proliferation;
KW   nervous system disorder; inflammatory disorder; cell differentiation;
KW   angiogenesis; stem cell growth factor; activin; inhibin; cartilage; burn;
KW   genetic disorder; bone regeneration; tendon; ligament; tissue repair;
KW   cytostatic; antirheumatic; antiarthritic; vulnerary; antiinflammatory;
KW   antibacterial; immunosuppressive; vasotropic; antiparkinsonian;
KW   neuroprotective; osteopathic; antidiabetic; antiasthmatic; antiallergic;
KW   immunostimulant; analgesic; gene therapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO200164834-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04926.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
PR   17-JUN-2000; 2000US-0597707.
PR   14-JUL-2000; 2000US-0616807.
PR   19-SEP-2000; 2000US-0664641.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI   Xue AJ, Yang Y, Wehrman T, Wang J, Ma Y, Wang D, Chen R, Xu C;
PI   Drmanac R;
XX
DR   WPI; 2001-589862/66.
DR   N-PSDB; AAS44894.
XX
PT   Novel polypeptides and nucleic acids obtained from cDNA libraries
PT   prepared from various human tissues, for diagnosis, treatment of
PT   cancer, neurological, inflammatory disorders and for use in arrays for
PT   detection -
XX
PS   Claim 10; Page 144; 153pp; English.
XX
CC   Sequences AAU27676-AAU28019 represent full-length polypeptides and
CC   contig polypeptides of the invention. The proteins and their associated
CC   DNA sequences are useful for the treatment, diagnosis and prevention of
CC   various types of disorder in a mammalian subject such as a human, dog,
CC   monkey, mouse, hamster or rat. The disorders include cancers such as
CC   leukaemia, lymphoma and neuroblastoma, autoimmune disorders such as
CC   multiple sclerosis, connective tissue disease, rheumatoid arthritis,
CC   diabetes mellitus, allergic rhinitis, asthma and eczema, nervous system
CC   disorders such as Parkinson's disease, Alzheimer's disease, Huntington's
CC   chorea, amyotrophic lateral sclerosis, spinal muscular atrophy and
CC   Wernicke disease, inflammatory disorders such as nephritis, Crohn's
CC   disease, ischaemia-reperfusion injury, shock, sepsis and inflammatory
CC   bowel disease. The sequences exhibit activity relating to angiogenesis,
CC   cell proliferation, cell differentiation, stem cell growth factor,
CC   activin or inhibin. Therefore, they can be used to manipulate stem cells
CC   in culture to give rise to neuroepithelial cells that can be used to
CC   augment or replace cells damaged by illness, accidental damage or genetic
CC   disorders. The sequences may also be used for regeneration of bone,
CC   cartilage, tendons and ligaments and in tissue repair and burn healing.
CC   Note: Some sequences for this patent did not form part of the printed
CC   specification, but were obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
```

```
SQ   Sequence   104 AA;


alignment_scores:
           Quality:  319.00            Length:     96
             Ratio:  4.833               Gaps:      1
 Percent Similarity:  68.750   Percent Identity: 68.750

alignment_block:
 US-09-664-641-10 x AAU27994   ..

 Align seg 1/1  to: AAU27994  from: 1  to: 104

    2833 TACATCACACCTGGAATCTGCCCAAGTCTTTCCACTATGAAGGCAATCGT 2882
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 TyrIleThrProGlyIleCysProSerLeuSerThrMetLysAlaIleVa 17

    2883 AGAGTGTGCAGGAGGAAAGGTGTTATCCAAGCAGCCATCTTTCCGGAAGC 2932
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      17 lGluCysAlaGlyGlyLysValLeuSerLysGlnProSerPheArgLysL 34

    2933 TCATGGAGCACAAGCAGAACTCG..................................... 2955
         |||||||||||||||||||||||
      34 euMetGluHisLysGlnAsnSerValGlyArgValGluSerAlaSerGlu 50

    2955 ..................................................... 2955

      51 SerValIleAlaLeuTyrHisIleGlnArgSerValThrGluLysLeuIl 67

    2956 .............AGTTTGTCGGAAATAATTTTAATATCCTGTGAAAATG 2992
                       |||||||||||||||||||||||||||||||||||||
      67 ePhePheSerLeuSerLeuSerGluIleIleLeuIleSerCysGluAsnA 84

    2993 ACCTTCATTTATGCCGAGAATATTTTGCCAGAGGCATA 3030
         |||||||||||||||||||||||||||||||||||||
      84 spLeuHisLeuCysArgGluTyrPheAlaArgGlyIle 96

seq_name: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAO05204

seq_documentation_block:
ID   AAO05204 standard; Protein; 143 AA.
XX
AC   AAO05204;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polypeptide SEQ ID NO 19096.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200164835-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04927.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   N-PSDB; AAI85135.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing
PT   diagnosing and treating e.g. leukaemia, inflammation and immune
PT   disorders -
XX
PS   Claim 20; SEQ ID NO 19096; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AAI79941-AA193841) and
CC   the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   143 AA;


alignment_scores:
           Quality:  294.50            Length:     69
             Ratio:  4.750               Gaps:      1
 Percent Similarity:  89.855   Percent Identity: 84.058

alignment_block:
 US-09-664-641-10 x AAO05204   ..

 Align seg 1/1  to: AAO05204  from: 1  to: 143

    1471 CTACAGCAGCAGCAGCAGGCCCAGCAGCAGCAGCAGCAGCACCCGGTTTT 1520
         :::|||||||||||||:::||||||:::|||||||||||||||||||||||
       1 ValGlnGlnGlnGlnArgAlaGlnGluGlnGlnGlnGlnHisProValLe 17

    1521 ACACCTTCAGCCCCAGCAGATAATGCAGCTCCAGCAGCAGCAGCAGCAGC 1570
         |||||||||||||||||||||||||||||||||||||||||||||||:::|
      17 uHisLeuGlnProGlnGlnIleMetGlnLeuGlnGlnGlnGlnGlnArgG 34

    1571 AGATCTCTCAGCAACCTTACCCCCAGCAGCCGCCGCATCCATTTTCACAG 1620
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      34 lnIleSerGlnGlnProTyrProGlnGlnProProHisProPheSerGln 50

    1621 CAGCAGCAGCAGCAACAGCAGCCACCACCA.........TCGCCTCAGCA 1661
         |||||||||||||||||||||       |||         |||   |||||
      51 GlnGlnGlnGlnGlnGlnAlaHisProHisGlnPheSerGlnGlnGl 67

    1662 GCATCAG 1668
         |   |||
      67 nLeuGln 69

seq_name: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY03183

seq_documentation_block:
ID   AAY03183 standard; protein; 1522 AA.
XX
AC   AAY03183;
XX
DT   16-JUN-1999  (first entry)
XX
DE   Topoisomerase II binding protein 2.
XX
KW   Topoisomerase II binding protein; TopBP; anticancer agent.
XX
OS   Homo sapiens.
XX
PN   JP11075856-A.
XX
PD   23-MAR-1999.
XX
PF   17-SEP-1997;   97JP-0251544.
XX
```

```
PR   17-SEP-1997;   97JP-0251544.
XX
PA   (CHUS ) CHUGAI PHARM CO LTD.
PA   (TSUR/) TSURUO T.
XX
DR   WPI; 1999-257704/22.
DR   N-PSDB; AAX28154.
XX
PT   New Topoisomerase II- binding protein - useful as an anticancer
PT   agent
XX
PS   Claim 7; Page 14-18; 28pp; Japanese.
XX
CC   This sequence represents the topoisomerase II binding protein (TopBP) of
CC   the invention. The TopBP protein is useful as an anticancer agent. TopBP
CC   can be used as the target molecule for anticancer agent.
XX
SQ   Sequence   1522 AA;


alignment_scores:
            Quality:  263.50              Length:   1031
              Ratio:  0.620                Gaps:     43
 Percent Similarity:  41.222   Percent Identity: 19.011

alignment_block:
 US-09-664-641-10 x AAY03183   ..

 Align seq 1/1  to: AAY03183  from: 1  to: 1522

      234 AATGCACTAGCCTCACACATAATCTCAGAGGATGGGGACAATCCAGAGGT 283
          |||:::   ::::::|||:::|||::: |||||
      140 AsnValSerValThrHisLeuIleAla.............GlyGluVa 151

      284 GGGAGAAGCCTCGGGAAGTC...........TTTGACTTACCTGTTGTAA 321
          ||||:::   :::   :::          :::    |||::::::
      151 lGlySerLysLysTyrLeuValAlaAlaAsnLeuLysLysProIleLeuL 168

      322 AGCCTTCTTGGGTGATTCTG.....................TCCGTT 347
            ||||||||||:::                          ::::::
      168 euProSerTrpIleLysThrLeuTrpGluLysSerGlnGluLysLysIle 184

      348 CAGTGTGGAACTCTTCTGCCAGTAAATGGTTTTTCTCCAGAATCATGTCA 397
                |||   :::   ::::::   |||        :::|||
      185 ThrArgTyrThrAspIleAsnMetGluAspPhe.....LysCysPr 198

      398 GATTTTTTTTGGAATCACTGCCTGCCTTTCTCAGGGTGTTGATACAAGCT 447
           ||||||:::|||        |||::::::
      198 oIlePheLeuGlyCysIleIleCysValThr.................. 208

      448 GGAGCTCTTTGTTGCAGTCTTCCAGAGCTCTCCCAAGGAGAGGTAGGGAA 497

      208 ..................................................  208

      498 GGGAGCTTGTCCAGCAGAAGTTGGGAAGCACAGAGATCATCTGCCTTCTT 547

      208 ..................................................  208

      548 CTGACCCGGTATTGATGCAGGCTGAGGCCTCTGTTGTAATGTGCTGGGTG 597
                            |||||||||    :::
      209 .....................GlyLeuCys.GlyLeu.....  213

      598 TCATCTGAAGACAGAAGTGCCCTGTGGGCTTTGGTTACGTTCTATGGGGG 647
               |||||:::   :::      |||::::::   :::||||
      214 ........AspArgLysGluValGlnGlnLeuThrValLysHisGlyGl 227

      648 AGATTGC......CAGCTAACCCTCAATAAGAAATGCACGCATTTGATTG 691
          |:::         ||||||   :::|||   :::|||||||||||||||||
      227 yGlnTyrMetGlyGlnLeuLysMetAsn...GluCysThrHisLeuIleV 243

      692 TTCCAGAGCCAAAGGGGGGGAAATACGAATGTGCTTTAAAGCGAGCAAGT 741
          ||  ||||||||||||  ||||||||||||||||  ||||||  :::
      243 alGlnGluProLysGlyGlnLysTyrGluCysAla...LysArgTrpAsn 258

      742 ATTAAAATTGTGACTCCTGACTGGGTTCTGGATTGCGTATCAGAGAAAAC 791
          :::      ||||||  :::|||  :::|||  :::::::::
      259 ValHisCysValThrThrGlnTrpPhePheAspSerIleGluLysGlyPh 275

      792 CAAAAAGGACGAAGCATTTTATCATCCTCGTCTGATTATTTATGAAGAGG 841
              :::||||||:::::::|||                          :
      275 eCysGlnAspGluSerIleTyr.........................L 283

      842 AAGAAGAGGAAGAGGAAGAGGAGGAGGAAGTAGAAAATGAGGAACAAGAT 891
          ::   |||  :::  |||  :::  :::  |||:::::
      283 ysThrGluProArgProGluAlaLysThrMetProAsnSerSerThrPro 299

      892 TCTCAGAATGAGGGTAGTACAGATGAGAAGTCAAGCCCTGCCAGCTCTCA 941
          :::::::::   ::::::   |||:::::::::
      300 ThrSerGlnIleAsnThrIleAspSerArgThr................. 310

      942 AGAAGGGTCTCCTTCAGGTGACCAGCAGTTTTCACCTAAATCCAACACTG 991
                                      :::|||      ||||||  :
      311 ..........................LeuSerAspValSerAsnIleS 318

      992 AAAAATCTAAAGGGGAATTAATGTTTGATGATTCTTCAGATTCATCACCG 1041
          :::::  :::::::::  :::  ::::::      :::|||
      318 erAsnIleAsnAlaSerCysValSerGluSerIleCysAsnSerLeuAsn 334

     1042 GAAAAACAGGAGAGAAATTTAAACTGGACCCCGGCCGAAGTCCCACAGTT 1091
          :::|||   |||  :::|||:::           ::::::  :::||
      335 SerLysLeuGluProThrLeuGlu...........AsnLeuGluAsnLe 347

     1092 AGCTGCAGCAAAACGCAGGCTGCCTCAGGGAAAGGAGCCTGGGTTG...A 1138
          |  ::::::      :::  |||:::            |||      |
      347 uAspValSerAlaPheGlnAlaProGluAspLeuLeuAspGlyCysArgI 364

     1139 TTAACTTGTGTGCCAATGTCCCACCCGTCCCAGGTAACATTTTGCCCCCT 1188
          ||  ||||||:::
      364 leTyrLeuCysGly................................... 368

     1189 GAGGTCCGGGGTAATTTAATGGCTGCTGGACAAAACCTCCAAAGTTCTGA 1238
                                    :::|||::::::|||::::::
      369 ...................PheSerGlyArgLysLeuAspLys..... 376

     1239 AAGATCAGAAATGATAGCTACCTGGAGTCCAGCTGTACGGACACTGAGGA 1288
                                          :::|||   |||  |
      377 ................................LeuArgArgLeuIleA 382

     1289 ATATTACTAATAATGCTGACATTCAGCAGATGAACCCGGCCATCAAATGTA 1338
          ||      :::::::::   ::::::|||:::|||      ::::::|||
      382 snSerGlyGlyGlyValArgPheAsnGlnLeuAsn......GluAspVal 396

     1339 GCACATATCTTACAGACTCTTTCAGCACCTACGAAAAATTTAGAACAGCA 1388
          :::|||::::::      ::::::           :::|||:::|||
      397 ThrHisValIle......ValGlyAspTyrAspAspGluLeuLysGlnPh 411

     1389 GGTGAATCACAGC...CAGCAGGGACATACAAATGCCAATGCAGTGCTGT 1435
              |||  |||  ::::::  |||:::  :::        |||
      411 eTrpAsnLysSerAlaHisArgProHisValValGlyAlaLysTrpLeu. 427

     1436 TTAGCCAAGTGAAAGTGACTCCAGAGACACACATGCTACAGCAGCAGCAG 1485
               ::::::       :::  :::||||||:::::::::
      428 ........LeuGluCysPheSerLysGlyTyrMetLeuSerGluGluPro 441

     1486 ......CAGGCCCAGCAGCAGCAGCAGCAGCACCCGGTTTTACACCTTCA 1529
               :::|||:::  |||      :::  ||||||  |||  ||
      442 TyrIleHisAlaAsnTyrGlnProValGluIleProValSerHis...Gl 457

     1530 GCCCCAGCAGATAATGCAGCTCCAGCAGCAGCAGCAGCAGATCTCTC 1579
          |||||::::::         |||  ::::::::::::::::::::|||:
      457 nProGluSerLysAlaAlaLeuLeuLysLysLysAsnSerSerPheSerL 474

     1580 AGCAACCTTACCCCCAGCAGCCCGCCGCAT..................... 1608
          :::::  :::      :::      |||
      474 ysLysAspPheAlaProSerGluLysHisGluGlnAlaAspGluAspLeu 490
```

```
1608 ..................................................... 1608

 491 LeuSerGlnTyrGluAsnGlySerSerThrValValGluAlaLysThrSe 507

1609 ..........CCATTTTCACAGCAGCAGCAGCAGCAA.............. 1635
                |||||:::::::::   :::   :::
 507 rGluAlaArgProPheAsnAspSerThrHisAlaGluProLeuAsnAspS 524

1636 ................CAGCAGCCACCACCATCGCCTCAGCAGCATCAG 1668
                     |||:::         |||      :::|||
 524 erThrHisIleSerLeuGlnGluGluAsnGlnSerSerValSerHisCys 540

1669 CTTTTTGGACATGATCCAGCAGTGGAG...ATTCCAGAAGAAGGCTTCTT 1715
     :::            |||   |||:::   |||   |||||||||:::::
 541 Val............ProAspValSerThrIleThrGluGluGlyLeuPh 553

1716 ATTGGGATGTGTGTTTGCAATTGCGGATTATCCAGAGCAGATGTCTGATA 1765
     :          |||   :::      :::   ::::::   :::
 553 eSerGlnLysSerPheLeuValLeuGlyPheSerAsnGluAsnGlu.... 568

1766 AGCAACTGCTGGCCACCTGGAAAAGGATAATCCAGGCACATGGCGGCACT 1815
                :::::      :::|||||||:::   :::::|||
 569 ..........SerAsnIleAlaAsnIleIleLysGluAsnAlaGlyLys 581

1816 GTTGACCCCACCTTCACGAGTCGATGCACGCAC.................. 1848
     :::    ::::::   ||||||   :::
 582 Ile...MetSerLeuLeuSerArgThrValAlaAspTyrAlaValValPr 597

1849 ....CTTCTCTGTGAGAGTCAAGTCAGCAGGCGCTATGCACAGGCAATAA 1894
         |||   ||||||   :::::::::   :::
 597 oLeuLeuGlyCysGluValGluAlaThrValGly................ 608

1895 GAGAAAGAAACAGATGTGTTACTGCACACTGGTTAAACACAGTCTTAAAA 1944
               :::   ||||||      ||||||   |||   :::
 609 ..........GluValValThrAsnThrTrpLeuValThrCysIleAsp 621

1945 AAAAAAAAAATGGTA...CCGCCGCACCGAGCCCTTCACTTCCCAGTGGC 1991
     :::   :::      |||      :::   |||      ||||||
 622 TyrGlnThrLeuPheAspProLysSerAsnProLeuPheThrProValPr 638

1992 CTTCCCACCAGGAGGAAAGCCATGTTCACAGCATATTATTTCTGTGACTG 2041
           |||      |||   ::::::   :::||||||   :::
 638 oValMetThrGlyMetThrProLeuGluAspCysValIleSerPheSerG 655

2042 GATTTGTTGATAGTGACAGAGATGACCTAAAATTAATGGCTTATTTGGCA 2091
        :::   :::::::::::::::|||   :::::::||   |||
 655 lnCysAlaGlyAlaGluLysGluSerLeuThrPheLeuAlaAsnLeuLeu 671

2092 GGTGCCAAATATACGGGTTATCTATGCCGC..................... 2121
     ||||||:::          |||:::   |||
 672 GlyAlaSerValGlnGluTyrPheValArgLysSerAsnAlaLysLysGl 688

2122 .......AGCAACACAGTCCTCATCTGTAAAGAACCAACTGGTTTAAAGT 2164
            ::::::|||   ||||||   ||||||      |||   ||||
 688 yMetPheAlaSerThrHisLeuIleLeuLysGluArgGlyGlySerLysT 705

2165 ATGAAAAAGCCCAAAGAGTGGAGGATACCCTGTGTCAACGCCCAGTGGCTT 2214
     |||||   ||||||:::|||:::::::|||:::|||:::      ||||||
 705 yrGluAlaAlaLysLysTrpAsnLeuProAlaValThrIleAlaTrpLeu 721

2215 ..................................GGCGACATTCTTCT 2228
                                        :::   :::|||::
 722 LeuGluThrAlaArgThrGlyLysArgAlaAspGluSerHisPheLeuIl 738

2229 GGGAAAC.......................................... 2235
     :   |||
 738 eGluAsnSerThrLysGluGluArgSerLeuGluThrGluIleThrAsnG 755

2235 ..................................................... 2235

 755 lyIleAsnLeuAsnSerAspThrAlaGluHisProGlyThrArgLeuGln 771

2236 ....................................................TT 2237

 772 ThrHisArgLysThrValValThrProLeuAspMetAsnArgPheGlnSe 788

2238 TGAGGCACTGAGGCAGATT......CAGTATAGTCGCTACACGGCATTCA 2281
       :::|||:::|||   :::      |||:::::::|||   :::|||   |
 788 rLysAlaPheArgAlaValValSerGlnHisAlaArgGlnValAlaAlaS 805

2282 GT.............................................. 2283
     ||
 805 erProAlaValGlyGlnProLeuGlnLysGluProSerLeuHisLeuAsp 821

2283 ..................................................... 2283

 822 ThrProSerLysPheLeuSerLysAspLysLeuPheLysProSerPheAs 838

2284 ......................CTGCAGGATCCATTTGCCCCTACCCAGC 2311
                           |||:::   |||      |||:::|||:
 838 pValLysAspAlaLeuAlaAlaLeuGluThrProGlyArgProSerGlnG 855

2312 AT.............................TTAGTTTTAAATCTT 2328
     ::                                :::|||   ||||||
 855 lnLysArgLysProSerThrProLeuSerGluValIleValLysAsnLeu 871

2329 TTAGATGCTTGGAGAGTTCCCTTAAAAGTGTCTGCAGAGTTGTTGATGAG 2378
                ::::::   |||      |||::::::      :::::
 872 ............GlnLeuAlaLeuAlaAsnSerSerArgAsnAlaValA. 884

2379 TATAAGACTACCTCCCAAACTGAAACAGAATGAAGTAGCTAATGTCCAGC 2428
     ::::         |||:::|||||:::
 884 aLeuSerAlaSerProGlnLeuLysGlu..................... 893

2429 CTTCTTCCCAAAGAGCCAGAATTGAAGACGTACCACCTCCCACTAAAAAG 2478
               :::::::::   |||:::::::|||   |||      |||
 894 ........AlaGlnSerGluLysGluGluAlaProLysProLeuHisLys 907

2479 CTAACTCCAGAATTGACCCCTTTTGTGCTTTTCACTGGATTCGAGCCTGT 2528
     ::::::      ::::::
 908 ValValValCysValSer............................... 913

2529 CCAGGTTCAACAGTATATTAAGAAGCTCTACATTCTTGGTGGAGAGGTTG 2578
          ::::::      ||||||         |||:::|||   :::|
 914 .......LysLysLeuSerLysLysGlnSerGluLeuAsnGlyIleAlaA 928

2579 CGGAGTCTGCA.....................CAGAAGTGCACACAC 2604
     ||:::   :::                     :::      ||||||
 928 laSerLeuGlyAlaAspTyrArgTrpSerPheAspGluThrValThrHis 944

2605 CTCATT............GCCAGCAAAGTGACTCGCACCGTGAAGTTCCT 2642
     :::|||            :::   ::::::|||      |||   ::
 945 PheIleTyrGlnGlyArgProAsnAspThrAsnArgGluTyrLysSerVa 961

2643 GACGGCGATTTCTGTCGTGAAGCACATAGTGACGCCAGAGTGGCTGGAAG 2692
     :          :::|||      ||||||||||:::   :::||||||   :
 961 lLysGluArgGlyVal......HisIleValSerGluHisTrpLeuLeuA 976

2693 AATGCTTCAGGTGTCAGAAGTTCATTGATGAGCAGAACTAC 2733
     ::|||   :::      |||   :::   |||:::   |||
 976 spCysAlaGlnGluCysLysHisLeuProGluSerLeuTyr 989
```

seq_name: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG23599

seq_documentation_block:

ID ABG23599 standard; Protein; 61 AA.
XX
AC ABG23599;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #23590.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX

```
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS87786.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 53958; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   61 AA;


alignment_scores:
            Quality:  249.00              Length:     84
              Ratio:  6.385                Gaps:      5
 Percent Similarity:  46.429   Percent Identity: 45.238

alignment_block:
 US-09-664-641-10/rev x ABG23599   ..

 Align seq 1/1  to: ABG23599  from: 1  to: 61

    1650 TGGTGGTGGCTGCTGTTGCTGCTGCTGCTGCTGTGAAAATGGATGCGGCG 1601
         ||||||||| ||||||||||||||||||||||||          |||
       7 TrpTrpTrp.CysCysCysCysCysCysCys.........CysCysC 20

    1600 GCTGCTGGGGGTAAGGTTGCTGAGAGATCTGCTGCTGCTGCTGCTGCTGG 1551
           |||           |||         ||||||||||||||||||
      20 ysCys...........CysCysCysCysCysCysCysCysCysCys 32

    1550 AGCTGCATTATCTGCTGGGGCTGAAGGTGTAAAACCGGGTGCTGCTGCTG 1501
           |||      |||           |||          ||||||||||
      33 CysCysCysCys...........CysCysCysCysCysCysCy 45

    1500 CTGCTGCTGGGCCTGCTGCTGCTGCTGTAGCATGTGTGTCTCTGGAGTCA 1451
```

```
         |||||||   :::|||||||||||||||||      |||
      45 sCysCysCysCysCysCysCysCysCysCysCys..... 57

    1450 CTTTCACTTGGCTAAACAGCACTGCATTGGCATTTGTATGTCCCTGCTGG 1401
                                              |||   ||||||
      58 .....CysCysCysTrp 61

seq_name: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG19008

seq_documentation_block:
ID   ABG19008 standard; Protein; 398 AA.
XX
AC   ABG19008;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #18999.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS87795.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 49367; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   398 AA;


alignment_scores:
            Quality:  248.00              Length:    157
```

```
            Ratio:  3.875                Gaps:      7
 Percent Similarity:  40.764   Percent Identity: 29.936

alignment_block:
 US-09-664-641-10/rev x ABG19008   ..

 Align seq 1/1  to: ABG19008  from: 1  to: 398

    1730 AACACACATCCCAATAAGAAGCCTTCTTCTGGAATCTCCACTGCT..... 1686
         :::///     ::::::::::|||   :::|:|:::::::::::
     138 SerThrSerAspSerGlnSerProCysGlyGlyValGlySerSerLeuVa 154

    1685 .....................................................G 1685
                                                             |
     154 lVal***AlaGluArgSerCysAlaGlnCysArgMetSerProArgLysG 171

    1684 GATCATGTCCAAAAAGCTGATGCTGCTGACGCGATGGTGGTGGCTGCTGT 1635
         ||  |||          ||||||                  ||||||
     171 lyHisCys...........CysCys.................CysCys 177

    1634 TGCTGCTGCTGCTGCTGTGAAAATGGATGCGGCGGCTGCTGGGGGTAAGG 1585
         ||||||||||||||||||  :::  |||      |||:::
     178 CysCysCysCysCysCysSerCysCys......CysTyrCysHisTy 192

    1584 TTGCTGAGAGATCTGCTGCTGCTGCTGCTGCTGGAGCTGCATTATCTGCT 1535
          |||         ||||||||||||||||||      |||      |||
     192 rCysCysCysTyrCysCysCysCysCysCysCysCysCysCysCys. 208

    1534 CGGGCTGAAGGTGTAAAACCGGGTGCTGCTGCTGCTGCTGCTGGGCCTGC 1485
               |||         ||||||||||||||||||   :::|||
     209 ..........CysCysCysCysCysCysCysCysCysCysCys 221

    1484 TGCTGCTGCTGTAGCATGTGTGTCTCTGGAGTCACTTTCACTTGGCTAAA 1435
         ||||||||||||      |||
     222 CysCysCysCysTyrCysCys........................... 228

    1434 CAGCACTGCATTGGCATTTGTATGTCCCTGCTGGCTGTGATTCACCTGCT 1385
                        |||  |||         :::  ||||
     229 ....................CysArgCys.........TyrCysCysC 235

    1384 GTTCTAAATTTTTCGTAGGTGCTGAAAGAGTCTGTAAGATATGTGCTACA 1335
         ||        :::      :::         |||:::  |||:::
     235 ysCysCysCysTyrCysCysCysCysCysCysArgCysCysCysCys 251

    1334 TTTGATGGCCGGTTCATCTGC 1314
               |||      |||
     252 CysCysCysArgCysCysCys 258

seq_name: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG23600

seq_documentation_block:
ID   ABG23600 standard; Protein; 87 AA.
XX
AC   ABG23600;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #23591.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS87787.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 20; SEQ ID No 53959; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   87 AA;


alignment_scores:
            Quality:  242.00                Length:    127
              Ratio:  4.566                   Gaps:      7
 Percent Similarity:  41.732   Percent Identity: 36.220

alignment_block:
 US-09-664-641-10/rev x ABG23600   ..

 Align seq 1/1  to: ABG23600  from: 1  to: 87

    1742 TCCGCAATTGCAAACACACATCCCAATAAGAAGCCTTCTTCTGGAATCTC 1693
         |||   |||||||:::      |||:::::::  ||||||:::
       5 SerLeuIleAlaSer......ProArgSerLeuProSerAlaProGluLe 19

    1692 CACTGCTGGA.................TCATGTCCA...AAAAGCTGAT 1664
          :::||||||                  ||||||   :::|||  |
      19 uValAlaGlyLeuSerHisGlyCysGluLeuCysProAspArgSerProC 36

    1663 GCTGCTGAGGCGATGGTGGTGGCTGCTGTTGCTGCTGCTGCTGCTGTGAA 1614
         |||||                 ||||||||||||||||||||||||
      36 ysCys.................CysCysCysCysCysCysCys... 45

    1613 AATGGATGCGGCGGCTGCTGGGGGTAAGGTTGCTGAGAGATCTGCTGCTG 1564
                                                 ||||||||
      46 .....................................CysCysCy 48

    1563 CTGCTGCTGCTGGAGCTGCATTATCTGCTGGGGCTGAAGGTGTAAAACCG 1514
         ||||||||||       |||      |||
      48 sCysCysCysCysCysCysCysCys..................... 57

    1513 GGTGCTGCTGCTGCTGCTGCTGGGCCTGCTGCTGCTGCTGCTGTAGCATGTGT 1464
          ||||||||||||||||  :::|||||||||||||||      |||
      58 ..CysCysCysCysCysCysCysCysCysCysCysCysCysCys 73
```

```
1463 GTCTCTGGAGTCACTTTCACTTGGCTAAACAGCACTGCATTGGCATTTGT 1414

  74 .............................................Cy 74

1413 ATGTCCCTGCTGGCTGTGATTCACCTGCTGT 1383
       |||   |||         |||   ||||||
  74 sCysCysCysCysCysPheCysCysCys 84
```

seq_name: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU07343

seq_documentation_block:

```
ID   AAU07343 standard; Protein; 1679 AA.
XX
AC   AAU07343;
XX
DT   04-DEC-2001  (first entry)
XX
DE   1-aminocyclopropane carboxylate (ACPC) synthase #12.
XX
KW   1-aminocyclopropane carboxylate synthase; ACPC synthase; brain trauma;
KW   excitatory neurotransmission; neurodegeneration; stroke; nerve damage;
KW   neurodegenerative disease; Alzheimer's disease; depression; epilepsy;
KW   alcohol abuse; cognitive function; memory; learning impairment; human.
XX
OS   Homo sapiens.
XX
PN   WO200168879-A2.
XX
PD   20-SEP-2001.
XX
PF   14-MAR-2001; 2001WO-EP02857.
XX
PR   14-MAR-2000; 2000US-0189086.
PR   05-APR-2000; 2000US-0194702.
XX
PA   (FARB ) BAYER AG.
XX
PI   Ramakrishnan S;
XX
DR   WPI; 2001-550286/61.
XX
PT   Isolated polynucleotide encoding a human
PT   1-aminocyclopropane-carboxylate (ACPC) synthase, useful for treating
PT   brain trauma and neurodegenerative disease (e.g. Alzheimer's disease,
PT   depression, epilepsy) -
XX
PS   Claim 1; Page 231-237; 242pp; English.
XX
CC   The invention relates to reagents and methods for regulating excitatory
CC   neurotransmission, and to prevent neurodegeneration. The method involves
CC   the use of an expression vector or a reagent that modulates the activity
CC   of a 1-aminocyclopropane-carboxylate (ACPC) synthase polypeptide. The
CC   reagent is useful for modulating the activity of an ACPC synthase in a
CC   disease such as stroke, a nerve damage or a neurodegenerative disease.
CC   The ACPC synthase polypeptide, polynucleotides and modulators are also
CC   useful for treating brain trauma and neurodegenerative disease (e.g.
CC   Alzheimer's disease, depression, epilepsy). The ACPC synthase modulators
CC   are also useful for treating alcohol abuse and improve cognitive function
CC   and memory of patients with learning impairment. The present sequence
CC   represents the amino acid sequence of human 1-aminocyclopropane-
CC   carboxylate (ACPC) synthase #12, used in the method of the invention.
XX
SQ   Sequence   1679 AA;
```

alignment_scores:

```
           Quality:  238.00            Length:    193
             Ratio:  2.833               Gaps:     10
 Percent Similarity:  43.523   Percent Identity: 30.052
```

alignment_block:

US-09-664-641-10/rev x AAU07343 ..

Align seg 1/1 to: AAU07343 from: 1 to: 1679

```
1823 GGGTCAACAGTGCCGCCATGTGCCTGGATTATCCTTTTCCAGGTGGCCAG 1774
     |||:::|||:::     |||:::
  25 GlyAlaThrThrCysGlyCysCys..................................... 32

1773 CAGTTGCTTATCAGACATCTGCTCTGGATAATCCGCAATTGCAAACACAC 1724
      :::|||   :::      |||:::|||
  33 .GlyCysCysGlyThrCysCysThrGly..................... 41

1723 ATCCCAATAAGAAGCCTTCTTCTGGAATCTCCACTGCTGGATCATGTCCA 1674
                                   |||:::|||:::|||
  42 ............................CysThrGlyGlyAlaCysGly 48

1673 AAAAGCTGATGCTGCTGAGGCGATGGTGGTGGCTGC.............. 1638
           ||||||   |||   :::||||||||||
  49 .........CysCysThrGlyGlyAlaGlyGlyCysThrCysGlyAlaAl 62

1637 ................TGTTGC...TGCTGCTGCTGCTGTGAAAATGGAT 1607
                     ||||||   ||||||||||||||||   ::::::|
  62 aCysCysCysCysGlyCysCysGlyCysCysCysCysCysThrAlaC 79

1606 GC............GGCGGCTGCTGGGGGTAAGGTTGCTGAGAGATCTGC 1569
     ||           |||||||||          :::|||          |||
  79 ysCysCysCysAlaGlyGlyCysThrThrThrAlaCysThrCysCysCys 95

1568 TGCTGCTGCTGCTGCTGGAGCTGCATTATCTGCTGGGGCTGAAGGTGTAA 1519
     :::|||||||||||||   :::|||      |||   |||      |||
  96 AlaCysCysCysCysGlyGlyCysThrThrCysCysGlyCysCysCysAl 112

1518 A.........ACCGGGTGCTGCTGCTGCTGCTGGGCCTGCTGCTGCT 1478
             |||||||||      |||||||||   :::||||||   |
 112 aCysThrGlyThrGlyCysThrGlyCysCysCysThrThrCysCysThrC 129

1477 GCTGTAGCATGTGTGTCTCTGGAGTCACTTTCACT.............. 1443
     ||   :::   |||   :::|||          |||
 129 ysGlyGlyAlaCysCysThrGlyGlyGlyCysThrGlyThrCysGlyGly 145

1442 .......................TGGCTAAACAGCACTGCATTGGCATT 1417
                                    :::::::|||:::   :::
 146 GlyAlaGlyAlaGlyCysThrGlyGlyAlaGlyAlaThrGlyThrThrCy 162

1416 TGTATGTCCCTGCTGGCTGTGATTCACCTGCTGTTCTAAATTTTTCGTAG 1367
      :::|||   |||            ||||||||||:::         ::::
 162 sAlaCysCysCys.........ThrThrCysCysThr...CysAlaAlaA 175

1366 GTGCTGAAAGAGTCTGTAAGATATGTGCT 1338
     ::|||      :::|||      ||||||
 175 laAlaGlyGlyAlaCysThrThrCysAla 184
```

seq_name: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB69280

seq_documentation_block:

```
ID   ABB69280 standard; Protein; 776 AA.
XX
AC   ABB69280;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 34632.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
```

```
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL13383.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 34632; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   776 AA;


alignment_scores:
            Quality:  225.50              Length:    859
              Ratio:  0.616                Gaps:     40
 Percent Similarity:  42.608   Percent Identity: 21.886

alignment_block:
 US-09-664-641-10 x ABB69280  ..

 Align seg 1/1  to: ABB69280  from: 1  to: 776

     778 GTATCAGACAAAACCAAAAAGGACGAAGCATTTTATCATCCTCGTCTGAT 827
         ::::::::::::   ||||||          |||:::   :::||||::
      15 LeuAlaGlnGlnGlnLysLysLeuProLeuAlaTyrGlnThrAsnLeuVa 31

     828 TATTTATGAAGAGGAAGAAGAGGAAGAGGAAGAGGAGGAGGAAGTAGAAA 877
         :   |||:::
      31 lAspTyrArgThrAla................................. 36

     878 ATGAGGAACAAGATTCTCAGAATGAGGGTAGTACAGATGAGAAGTCAAGC 927
                    :::   :::::::::      :::::::::   :::
      37 ..............AlaTyrSerGlnAlaLeuTyrGlnGlnSerProThr 48

     928 CCTGCCAGCTCTCAAGAAGGGTCTCCTTCAGGTGACCAGCAGTTTTCACC 977
         :::||||||:::   |||:::|||   :::      ::::::::::||
      49 AlaThrSerSerSerGlyGlyGlyProLeuSerProIleAspMetGlnPr 65

     978 TAAATCC.........AACACTGAAAAATCTAAAGGGGAATTAATGTTTG 1018
         |:::|||         |||    |||      |||          :
      65 oGlnSerLysHisTyrAsnLeuMetLysHisGlyGlyGlyThr......S 80

    1019 ATGATTCTTCAGATTCATCACCGGAAAAACAGGAG.............. 1053
         :::::||||||   |||||||||      |||:::
      80 erSerSerSerHisSerSerProTyrHisGlnSerTyrSerSerSerGly 96

    1054 ..............AGAAATTTAAACTGGACCCCGGCCGAAGTCCCACA 1088
                       :::::::|||      :::||||:::|||
      97 GlyThrAlaAlaGlyGluGlnLeuTyrGlnSerProThrGluArgThrTy 113

    1089 GTTAGCTGCAGCAAAACGCAGGCTGCCTCAGGGAAAGGAGCCTGGGTTGA 1138
         ||||||||||||   |||      :::
     113 rLeuAlaAlaAlaGlyArgLeuHisAlaAsn.................. 123

    1139 TTAACTTGTGTGCCAATGTCCCACCCGTCCCAGGTAACATTTTGCCCCCT 1188
            |||      |||:::      |||:::   :::
     124 ..AsnAlaMetAlaGlyHisHisProIleSerAla.............. 134

    1189 GAGGTCCGGGGTAATTTAATGGCTGCTGGACAAAACCTCCAAAGTTCTGA 1238
                      |||::::::      |||:::||||||::::::::
     135 .............LeuGlnSerGlnTyrGlnGlnLeuGlnAlaAlaLy 146

    1239 AAGATCAGAAATGATAGCTACCTGGAGTCCAGCTGTACGG.......... 1278
         :   :::   ::::::||||||   |||   |||::::::
     146 sMetGlnAlaGlnMetAlaThrGlnSerGluAlaAlaGlnGlnGlnGlnA 163

    1279 ........ACACTGAGGAATATTACTAATAATGCTGACATTCAGCAGATG 1320
                 :::||||||                              |||   |||
     163 rgThrPheAlaLeuArg.......................GlnAlaMet 171

    1321 AACCGGCCATCAAATGTAGCACATATCTTACAG.................. 1353
         |||   |||:::|||      |||:::   |||
     172 AsnProProThrAsnHisTyrHisMetSerGlnSerSerSerMetValSe 188

    1354 .ACTCTTTCAGCACCTACGAAAAATTTAGAACAGCAGGTGAATCACAGCC 1402
          |||::::::|||   :::::::::   :::||||||
     188 rThrMetThrAlaLeuAsnGlnGlnGlnGlnGlnGln.............. 200

    1403 AGCAGGGACATACAAATGCCAATGCAGTGCTGTTTAGCCAAGTGAAAGTG 1452

     200 ..................................................... 200

    1453 ACTCCAGAGACACACATGCTACAGCAGCAGCAGCAGGCCCAGCAGCAGCA 1502
                              ||||||||||||||   ||||||||||
     201 ....................GlnGlnGlnGlnGlnGlnGlnGlnGlnGl 210

    1503 GCAGCAGCACCCGGTTTTACACCTT.................CAGCCCC 1534
         ||||:::   |||         |||                 ||||||
     210 nGlnArgAlaProProSerSerLeuAsnLeuGlnAsnGlnTyrGlnProA 227

    1535 AGCAG.........ATAATGCAGCTCCAGCAGCAGCAGCAGCAGCAGATC 1575
           |||           :::|||   ||||||||||||||||||||||
     227 laGlnGlyProLeuLysLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 243

    1576 TCTCAGCAACCT.....................................TA 1589
         ||||||||||||
     244 SerGlnGlnProAlaMetProLysHisTyrGlnGluGlnLeuTyrAlaGl 260

    1590 CCCCCAGCAGCCGCCGCATCCATTTTCACAGCAGCAGCAGCAGCAACAGC 1639
         ||||||      :::   :::   ||||||||||||||||||||||||
     260 nGlnGlnGlnLeuGlnGlnGlnGlnLeuHisGlnGlnGlnGlnGlnGlnG 277

    1640 AGCCACCACCATCGCCTCAGCAGCATCAGCTTTTTGGACATGATCCAGCA 1689
         ||         ::   |||||||||:::          |||
     277 lnLeuGlnGlnGlnGlnGlnGlnHisArg.........HisLysAsnAsp 290

    1690 GTGGACATTCCAGAAGAAGGCTTCTTATTGGGATGTGTGTTTGCAATTGC 1739
         ::::::   |||      |||      |||   |||:::
     291 LeuGlnThrPro......GlySerGluHisGlyThrValTyrIleGlnGl 305

    1740 GGATTATCCAGAGCAGATGTCTGATAAGCAACTGCTGGCCACCTGGAAAA 1789
          ::::::|||::::::::   ::::::
     305 nAsnHisProAsnHisValValAsnGln..................... 314

    1790 GGATAATCCAGGCACATGGCGGCACTGTTGACCCCACCTTCACGAGTCGA 1839

     314 ..................................................... 314

    1840 TGCACGCACCTTCTCTGTGAGAGTCAAGTCAGCAGCGCGTATGCACAGGC 1889
                   |||::::::|||:::|||::::::      :::||
     315 ............AlaCysGlnThrGlnIleSerAlaValLysProLysAl 327

    1890 AATA.........AGAGAAAGAAAGAGATGTGTTACTGCACACTGGTTAA 1930
         |           :::|||   :::         |||           :
     327 aThrProSerSerGluGluSerSerSerAsnSerThrLysSerProSerH 344

    1931 ACACAGTCTTAAAAAAAAAAAAAAAGTGGTACCGCCGCACCGAGCCCTTCAC 1980
         :::::   |||   :::||||||   :::         :::||||||
```

```
 344 isAlaProLeuAspArgLysLysSerAlaGlySerIleGlnAlaLeuLys 360

1981 TTCCCAGTGGCCTTCCCACCAGGAGGAAAGCCATGTTCACAGCATATTAT 2030
     |||::::::      |||   :::   |||                  ::
 361 SerProIleThrLysArgProProSerThrPro.............Va 372

2031 TTCTGTGACTGGATTTGTTGATAGTGAC...AGAGATGACCTAAAATTAA 2077
     ::::::::::|||:::::   ::::::      |||   |||||||:::
 372 lThrLeuSerGlyTrpLeuHisLysGlnGlySerAspGlyLeuLysVal. 388

2078 TGG..................................................... 2080
     |||
 389 TrpArgLysArgTrpPheValLeuAlaGluTyrCysLeuTyrTyrTyrLy 405

2081 ..................CTTATTTGGCAGGTG......CCAAATATAC 2105
                       |||:::   :::|||      |||:::   |
 405 sGlyProGluGluGluLysLeuLeuGlySerValLeuLeuProSerTyrA 422

2106 GGGTT.................................................. 2110
     |||||
 422 rgValSerAlaCysLeuProGluAspLysIleTyrArgLysPheAlaPhe 438

2111 .....................ATCTATGCCGCAGCAACACAGTCCTC 2136
                          |||      |||:::::::::||
 439 LysCysGluHisGlnAsnMetArgThrTyrTrpLeuAlaAlaAspAsnSe 455

2137 ATCTGTAAAGAACCAACTGGTTTAAAGT.................ATGA 2168
     |:::::::      |||   |||   :::                |||:
 455 rGluAlaMetMetGlnTrpValArgAlaLeuAlaAlaAlaSerLeuMetG 472

2169 AAAAGCCAAAGAGTGGAGGATACCCTGTGTCAACGCCCAGTGGCTTGGCG 2218
     ::   |||:::|||||||            |||:::   |||:::||| ::
 472 lnAlaProSerSerGlyGluSerGluProSerValAsnSerSerLeu.As 488

2219 ACATTCTTCTGGCAAACTTTCAGGCACTGAGGCAGATTCAGTATAGTCGC 2268
     :||||||                  |||   ::::::||||||      ::::
 488 nHisSer..............GlyGluAsnSerAspSer......GlyI 498

2269 TACACGGCATTCAGTCT...GCAGGATCCATTTGCCCCTACCCAGCATTT 2315
     ::|||   :::|||||||      |||:::      |||   |||:::
 498 leHisThrLeuGlnSerHisProGlyLysGlyGlnProThrProSer... 513

2316 AGTTTTAAATCTTTTAGATGCTTGGAGAGTTCCCTTAAAAGTGTCTGCAG 2365
               |||   :::      :::|||:::      ||| ::::::
 514 .........SerGluAsnThrGlySerSerGlyGlyGlySer.GlySerG 527

2366 AGTTGTTGATGAGTATAAGACTACCTCCCAAACTGAAACAG...AATGAA 2412
          :::      :::   |||||||||   ::::::   |||:::
 527 lyGlnProLeuTyrAlaAsnAlaProProLysProArgArgIleAsnAsp 543

2413 GTAGCTAATGTCCAGCCTTCTTCCCAAAGAGCCAGAATTGAAGACGTACC 2462
     :::      :::|||||||   ::::::   :::   ::::::
 544 GlyGlyTyrSerSerProSerProGluHisAsnGluGlnGlnGlnGlnHi 560

2463 ACCTCCCACTAAAAAGCTAACTCCAGAATTGACCCCTTTTGTGCTTTTCA 2512
     |||      :::   ::::::|||
 560 sProSerArgArgLeuMetSerPro..................... 568

2513 CTGGATTCGAGCCTGTCCAGGTTCAACAGTATATTAAGAAGCTCTACATT 2562
                    |||   |||   ::::::   :::
 569 ................ThrGlnGlnLeuTyrGlnGlnGlnGlnHisGln 579

2563 CTTGGTGGAGAGGTTGCGGAGTCTGCACAGAAGTGCACACACCTCATTGC 2612
     :::   :::      ::::::      :::      |||   |||
 580 ArgSerGlnGlnGlnGlnGlnGlnGlnProGlnGlnHisHisAlaIleTy 596

2613 CAGCAAAGTGACTCGCACCGTGAAGTTCCTGACGGCGATTTCTGTCGTGA 2662
     :::      |||||||||
 596 rAsp......ThrArgThr.........................G 601

2663 AGCACATAGTGACGCCAGAGTGGCTGGAAGAATGCTTCAGGTGTCAGAAG 2712
     |||:::   |||        |||:::::::::
 601 lyHisValSerThrAlaLeuGlnLeuGlnGlnAla.............. 612

2713 TTCATTGATGAGCAGAACTACATTCTCCGAGATGCTGAGGCAGAAGTA.. 2760
               :::|||:::|||   |||         |||||||:::
 613 .........GlnGlnGlnTyrSerLeuAspHisLeuGluAlaGlnPheGl 626

2761 .CTTTTCTCTTTCAGCTTGGAAGAATCCTTAAAACGGGCACACGTTTCTC 2809
              :::::::::::::|||||||::::::   |||
 626 nGlnGlnGlnLeuAspMetGluGluGlnIleAlaArg.............. 638

2810 CACTCTTTAAGGCAAAATATTTTTACATCACACCTGGAATCTGCCCAAGT 2859

 638 ..................................................... 638

2860 CTTTCCACTATGAAGGCAATCGTAGAGTGTGCAGGAGGAAAGGTGTTATC 2909
     |||:::   :::::|||      |||      |||            ::
 639 LeuGlnGlnGlnArgAlaAlaGluGluIleTyrGly............Gl 651

2910 CAAGCAGCCATCTTTCCGGAAGCTCATGGAGCACAAGCAGAACTCGAGTT 2959
     :::::::      :::   |||||||||:::::::::|||:::   :::
 651 uArgGluMetTyrMetAlaLysLeuMetGlnGlnArgGlnGlyProAsn. 667

2960 TGTCGGAAATAATTTTAATATCCTGTGAAAATGACCTTCATTTATGCCGA 3009
                    ||||||::::::::   ::::::   |||||
 668 GlyThr............TyrProThrGlnGlnGlnLeuLeuGlnAlaGl 680

3010 GAATATTTTGCCAGAGGCA 3028
     |:::      |||:::|||
 680 uArgArgThrProAspAla 686
```

seq_name: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB63516

seq_documentation_block:

ID ABB63516 standard; Protein; 1186 AA.
XX
AC ABB63516;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila me anogaster polypeptide SEQ ID NO 17340.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL07619.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 17340; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1186 AA;

alignment_scores:
Quality: 220.50 Length: 650
Ratio: 0.817 Gaps: 26
Percent Similarity: 41.538 Percent Identity: 21.538

alignment_block:
US-09-664-641-10 x ABB63516 ..

Align seq 1/1 to: ABB63516 from: 1 to: 1186

```
 865 GAGGAAGTAGAAAATGAGGAACAAGATTCTCAGAATGAGGGTAGTACAGA 914
     |||:::|||||| ::::::||| |||:::
 449 GluGlnValGluAlaHisLysGln......LeuIleGluAlaLeuProGl 463

 915 TGAGAAGTCAAGCCCTGCCAGCTCTCAAGAAGGGTCTCCTTCAGGTGACC 964
     ::: ::::::|||::::::::::::::: ||||||
 463 yAspLeuAsnThrProValThrThrSerSerAspIleProSer..... 477

 965 AGCAGTTTTCACCTAAATCCAACACTGAAAAATCT......AAAGGGGAA 1008
     |||:::||| ::: :::||| |||:::
 478 ..PhePheGlyProThrThrValValGluProProIleThrGlySer 493

1009 TTAATGTTTGATGAT..... 1023
     ::: :::|||
 494 IleGluSerGluAspLeuSerLeuGluProGlnValIleSerValAlaSe 510

1024 .......TCTTCAGATTCATCACCGGAAAAACAGCAGAGAAATTTAAACT 1066
     ||| |||||| |||:::|||||::
 510 rProValLeuSerHisCysSerProLeuLysGluGluArgSer..... 524

1067 GGACCCCGGCCGAAGTCCCACAGTTAGCTGCAGCAAAACGCAGGCTGCCT 1116
     |||||| ||| |||||| :::
 525 ..ThrPro.....ProAlaLeuAlaIleVal..... 532

1117 CAGGGAAAGGAGCCTGGCTTGATTAACTTGTGTGCCAATGTCCCACCCGT 1166
     |||||| ::: ||| |||
 533 ......LysGluGluSerSerAsnAsnSerCys..... 541

1167 CCCAGGTAACATTTTGCCCCCTGAGGTCCGGGGTAATTTAATGGCTGCTG 1216
     |||::: ||| ::: ::::::|||
 542 .......AsnMetTyrProGlnHisAsnAsnAsnAsn..... 551

1217 GACAAAACCTCCAAAGTTCTGAAAGATCAGAAATGATAGCTACCTGGAGT 1266
     ::::::::: :::||| |||
 552 .....AsnAsnAsnAsnAsnThrThrSerSer 560

1267 CCAGCTGTACGGACACTGAGGAATATTACTAATAATGCTGACATTCAGCA 1316
     :::::: ||| :::::: |||:::::::::::: :::
 561 SerThrThrThrThrSerLysGlnThrThrSerGluSerAsnThrGluCy 577

1317 GATGAACCGGCCATCAAATGTAGCACATATCTTACAGACTCTTTCAGCAC 1366
     :::::: |||:::||| :::::: |||
 577 sValGlySerProGlyAsnHisThrGlnSerHisGln..... 589

1367 CTACGAAAAATTTAGAACAGCAGGTGAATCACAGCCAGCAGGGACATACA 1416
     ::: :::||||||:::::|||:::::: ::: :::
 590 .......GlnGlnGlnGlnGlnLeuGlnHisAsnAsnThrSerSerSer 603

1417 AATGCCAATGCAGTGCTGTTTAGCCAAGTGAAAGTGACTCCAGAGACACA 1466
     |||:::|||::: ||
 604 AsnSerAsnCys.....HisHi 609

1467 CATGCTACAGCAGCAGCAGCAGGCCCAGCAGCAGCAG..... 1503
     | |||||||||||||| ||||||||||
```

```
 609 sSerHisGlnGlnGlnGlnGlnGlnGlnGlnGlnHisMetSerSerP 626

1504 .....CAGCAGCACCCGGTTTTACACCTTCAGCCCCAG... 1536
     ||||||||| :::|||||| ||| |||
 626 roGlnGlnGlnTyrGlnGlnHisGlnIleLeuHisGlnGlnGlnGlnLeu 642

1537 .....CAGAT 1541
     |||:
 643 ProAsnHisHisHisHisHisHisHisHisHisHisHisSerGlnLe 659

1542 AATGCAGCTCCAGCAGCAGCAGCAGCAGCAGATCTCTCAGCAACCTTACC 1591
     ::::|||||||||||||||||||||||||||| :::|||||| :::
 659 uGlnGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnHisG 676

1592 CCCAGCAGCCGCCGCATCCATTTTCACAGCAGCAGCAGCAGCAACAGCAG 1641
     ||||||||| ||| :::||| |||:::||||||||||||
 676 lnGlnGlnProLeuHisGlnGlnGlnGlnLeuGlnHisGlnGlnGlnGln 692

1642 CCACCACCATCGCCTCAGCAGCATCAGCTTTTTGGACATGATCCAGCAGT 1691
     |||:::|||||| ::::::
 693 .....GlnHisHisGlnGlnHisTyrGlnGlnHisAlaIl 704

1692 GGAGATTCCAGAAGAAGGCTTCTTATTGGGATGTGTGTTTGCAATTGCGG 1741
     ::::
 704 eGln..... 705

1742 ATTATCCAGAGCAGATGTCTGATAAGCAACTG...CTGGCCACCTGGAAA 1788
     :::|||:::|||:::::|||  :::::::::
 706 ........HisGlnGlnSerGlnGlnGlnAlaSerLysIleSerTyrArg 719

1789 AGGATAATCCAGGCACATGGCGGCACTGTTGACCCCACCTTCACGAGTCG 1838
     |||::: ::: |||:::::::::::::: ::: ::::::
 720 GlyIlePheThrThrThrGlyAsnAlaMetAsnAlaAlaAlaAlaAlaAl 736

1839 ATGCACGCACCTTCTCTGTGAGAGTCAAGTCAGCAGCGCGTATGCACAGG 1888
     :::::: ::: ||| :::::: ::: |||
 736 aAlaAlaAlaAlaGlnGlnGlnHisGlnGlnGlnGlnGlnHisGlnGlnG 753

1889 CAATAAGAGAAAGAAAGAGATGTGTTACTGCACACTGGTTAAACACAGTC 1938
     ::: ::: ::: ||| |||
 753 lnLeuProSerProGlnLeuGlyValLeuAla..... 763

1939 TTAAAAAAAAAAAAAAATGGTACCGCCGCACCGAGCCCTTCACTTCCCAGT 1988
     ||| |||||| ::::::|||
 764 .........GlyProMetSerProProSerAsnSerLeuGlyAsnSerTr 777

1989 GGCCTTCCCACCAGGAGGAAAGCCATGTTCACAG..... 2022
     :::::::||| ||| |||
 777 pGlyLeuProSerProAspLysThrMetPheGlnProProLeuPheSerL 794

2023 ........CATATTATTTCTGTG..... 2037
     ||| ::::::
 794 euProAlaHisTyrAlaThrMetGlnGlnGlnGlnGlnGlnGlnGlnGln 810

2038 .....ACTGCATTTGTTGATAGTGA 2057
     :::||| ::: |||
 811 GlnGlnGlnGlnGlnGlnGlnGlnAlaAlaGlyAlaAlaProSerPr 827

2058 CAGAGATGACCTAAAATTAATGGCTTATTTGGCAGGT......GCCAAAT 2101
     |||||| ::: ||| |||::: |||:::
 827 oTyrAspAspGlyArgAlaAlaAlaAlaAlaAlaAlaGlnHisAlaGluL 844

2102 ATACGGGTTATCTATGCCGCAGCAACACAGTCCTCATCTGTAAAGAACCA 2151
     ||| ||| ::: ||| |||::: |||:::|||
 844 euLeuGly...LeuThrMetAspCysThrProLeuLeuLeuLysGlnPro 859

2152 ACTGGTTTAAAGTATGAAAAAGCCAAAGAGTGGAGGATACCCTGTGTCAA 2201
     :::||| |||::: :::::
 860 ProPro...SerTyrAlaGlyAlaSer.....AlaGl 869

2202 CGCCCAGTGGCTTGGCGACATTCTTCTGGGAAACTTTGAGGCACTGAGGC 2251
     : ||||||||||::: :::::: ::: |||:::|
 869 yPheProGlyLeuGlyAspLeuHisSerSerHisGluGlnGlnLeuGlnG 886
```

```
 2252 AGATTCAGTATAGTCGCTACACGGCATTCAGTCTGCAGGATCCATTTGCC 2301
        ||   ||||||
  886 lnGlnGlnTyr..................................... 889

 2302 CCTACCCAGCATTTAGTTTTAAATCTTTTAGATGCTTGGAGAGTTCCCTT 2351

  889 ..................................................... 889

 2352 AAAAGTGTCTGCAGAGTTGTTGATGAGTATAAGACTACCTCCCAAACTG. 2400
                                  :::|||      ||||||
  890 ........................ValArgSerGlnProLysTyrG 897

 2401 ...........................AAACAGAATGAAGTAGCT 2418
                                  :::|||:::::::|||
  897 lnTrpLeuAspSerProAlaAspTyrAlaGlnGlnGlnGlnValGln 913

 2419 AATGTCCAGCCTTCTTCCCAAAGAGCCAGAATTGAAGACGTACCACCTCC 2468
      :::||||||   ::::::|||:::  :::          :::|||   ||
  914 GlnValGlnGlnGlnGlnGlnGlnGlnThrLeuValLeuProGlyPr 930

 2469 CACTAAAAAGCTAACT................................ 2484
      ||||::::::   :::
  930 oThrSerSerAlaSerSerSerAsnAlaAlaLeuGlyValLeuIleProL 947

 2485 ..............CCAGAATTGACCCCTTTTGTGCTTTTCACTGGATTC 2520
                     |||::::::   |||            |||||||::
  947 ysGlnGluAsnTyrProAspMetGlnProSerSerAsnGlyThrGlyTyr 963
```

seq_name: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG27446

seq_documentation_block:

ID ABG27446 standard; Protein; 89 AA.
XX
AC ABG27446;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #27437.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS91633.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 57805; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 89 AA;

alignment_scores:

Quality:	220.00	Length:	101
Ratio:	5.366	Gaps:	5
Percent Similarity:	40.594	Percent Identity:	35.644

alignment_block:
US-09-664-641-10/rev x ABG27446 ..

Align seg 1/1 to: ABG27446 from: 1 to: 89

```
 1700 GGAATCTCCACTGCTGGATCATGTCCAAAAAGCTGATGCTGCTGAGGCGA 1651
      |||   |||:::|||:::   :::          :::   ||||||
   24 GlyCysSerValAlaAlaProAlaAlaAlaAlaThrCysCys..... 37

 1650 TGGTGGTGGCTGCTGTTGCTGCTGCTGCTGCTGTGAAAATGGATGCGGCG 1601
           ||||||||||||||||||||||||
   38 .........CysCysCysCysCysCysCys.............. 45

 1600 GCTGCTGGGGGTAAGGTTGCTGAGAGATCTGCTGCTGCTGCTGCTGCTGG 1551
                                  ||||||||||||||||||
   46 ............................CysCysCysCysCys... 51

 1550 AGCTGCATTATCTGCTGGGGCTGAAGGTGTAAAACCGGGTGCTGCTGCTG 1501
        |||      |||                            ||||||||||
   52 ...CysCysCysCys.......................CysCysCysCy 59

 1500 CTGCTGCTGGGCCTGCTGCTGCTGCTGTAGCATGTGTGTCTCTGGAGTCA 1451
      |||||||   :::|||||||||||||||      |||
   59 sCysCysCysCysCysCysCysCysCysCysCys.............. 71

 1450 CTTTCACTTGGCTAAACAGCACTGCATTGGCATTTGTATGTCCCTGCTGG 1401
                                                 ||||||
   72 ............................................CysTrp 73

 1400 CTG 1398
      |||
   74 Leu 74
```

seq_name: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY60558

seq_documentation_block:

ID AAY60558 standard; Protein; 169 AA.
XX
AC AAY60558;
XX
DT 31-JAN-2000 (first entry)
XX
DE Human normal bladder tissue EST encoded protein 230.
XX
KW Human; bladder; treatment; EST; expressed sequence tag; cytostatic;
KW cancer; gene therapy.
XX
OS Homo sapiens.

```
XX
PN   DE19818620-A1.
XX
PD   28-OCT-1999.
XX
PF   21-APR-1998;   98DE-1018620.
XX
PR   21-APR-1998;   98DE-1018620.
XX
PA   (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI   Rosenthal A,  Specht T,  Hinzmann B,  Schmitt A,  Pilarsky C,  Dahl E;
XX
DR   WPI; 1999-602416/52.
DR   N-PSDB; AAZ42235.
XX
PT   New polypeptides and their nucleic acids, useful for treatment of
PT   bladder tumour and identification of therapeutic agents -
XX
PS   Claim 23; Page 338; 366pp; German.
XX
CC   This invention describes novel polypeptide fragment sequences (I) and
CC   their encoding nucleic acids (II) which are highly expressed in normal
CC   bladder tissue and have cytostatic activity. (II) are used for
CC   recombinant expression of (I) and to isolate complete genes. (I) are
CC   used to identify agents suitable for the treatment of bladder tumours,
CC   to directly treat this form of cancer (including expression from gene
CC   therapy vectors), or are used in a preparation for cancer treatment. (I)
CC   is also used for the generation of specific antibodies. (II) are
CC   identified by assembling ESTs (expressed sequence tags) from a
CC   particular tissue type before comparison of expression patterns. This
CC   allows a significantly longer fragment of the gene to be revealed, and
CC   therefore reduces the number of failures because of ESTs from different
CC   libraries representing different parts of the same unknown gene
CC   distorting the estimated frequency of occurrence in a particular tissue.
CC   AAY60329-Y60591 represent protein fragments encoded by the human normal
CC   bladder tissue cDNA library derived EST fragments represented in
CC   AAZ42122-Z42248.
XX
SQ   Sequence   169 AA;


alignment_scores:
           Quality:  216.00            Length:     86
             Ratio:  4.800               Gaps:      7
 Percent Similarity:  52.326   Percent Identity: 43.023

alignment_block:
 US-09-664-641-10/rev x AAY60558   ..

 Align seg 1/1  to: AAY60558  from: 1  to: 169

    1643 GGCTGCTGTTGCTGCTGC..............TGCTGCTGTGAAAATGG 1609
         |||||||||   ||||||                ::::::::
      58 GlyCysCysGlyCysCysGlyGlyProProLeuThrProTrpGlnArgAl 74

    1608 ATGCGGCGGC..............TGCTGGGGGTAAGGTTGCTGAGAGA 1574
         :|||||||||             ||||||      :::|||   :::
      74 aCysGlyGlyAspCysTrpSerSerCysTrp......SerCysSerAsnC 89

    1573 TCTGCTGCTGC...TGCTGCTGCTGGAGCTGCATTATC......TGCTGG 1533
           |||||||||   ||||||||||||||||||            ||||||
      89 ysCysCysCysAsnCysCysCysTrpSerCysCysCysCysAsnCysTrp 105

    1532 GGCTGAAGGTGTAAAACCGGGTGCTGCTGCTGCTGCTGCTGGGCCTGCTG 1483
         :::                  ||||||||||||||||||:::||||||
       106 Ser.....................CysCysCysCysCysTrpSerCysCy 115

    1482 CTGCTGCTGTAGCATGTGTGTCTCTGGAGTCACTTTCACTTGGCTAAACA 1433
         |||||||||                          |||||||||
       115 sCysCysCys.............................TrpLeuAsnM 122

    1432 GCACTGCA 1425
```

```
            :::|||
       122 etValAla 124

seq_name: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB61245

seq_documentation_block:
ID   ABB61245 standard; Protein; 517 AA.
XX
AC   ABB61245;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 10527.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL05348.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 10527; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   517 AA;


alignment_scores:
           Quality:  216.00            Length:    335
             Ratio:  1.278               Gaps:     10
 Percent Similarity:  50.448   Percent Identity: 27.463

alignment_block:
 US-09-664-641-10 x ABB61245   ..

 Align seg 1/1  to: ABB61245  from: 1  to: 517

     768 TCTGGATTGCGTATCAGAGAAAACCAAAAAGGACGAAGCATTTTATCATC 817
         :::|||   ||||||      :::      |||:::|||   :::   ::
     185 AsnGlyArgArgIleIleIleGlnAlaThrGlyAsnSerSerValLeuAl 201

     818 CTCGTCTGATTATTTATGAAGAGGAAGAAGAGGAAGAGGAAGAGGAGGAG 867
         ::::||||||  |||   ||||||:::|||||||||||||:::   |||
     201 aAlaSerAsp.IleLysGluGluAspGluGluGluGluAspProGluLeu 217
```

```
 868 GAAGTAGAAAATGAGGAACAAGATTCTCAGAATGAGGGTAGTACAGATGA 917
     :::   ||||||      ||||||   :::::::|||
 218 AsnAspGluAsn...LeuGlnAspIleGluGluGlu..............  228

 918 GAAGTCAAGCCCTGCCAGCTCTCAAGAAGGGTCTCCTTCAGGTGACCAGC 967
           :::   |||::::::   :::   :::|||:::          :
 229 .......AlaGlnAlaAlaGluAlaAspGluGluProGluThrThrAlaH 243

 968 AGTTTTCACCTAAATCCAACACTGAAAAATCTAAAGGGGAATTAATGTTT 1017
     :::::|||   :::   :::   |||   ::::::   :::   |||
 243 isLeuSerThrGlnValLysLeuGluLeuGluGluAspAspGluMetPro 259

1018 GAT............GATTCTTCAGATTCATCACCGGAAAAACAGGAGAG 1055
     |||           ::::::::::|||::::::   |||   |||:::
 260 AspGluGluGluGluGluAspGluAspAspAspGluGluLeuGlnHisGl 276

1056 AAATTTAAACTGGACCCCGGCCGAAGTCCCACAGTTAGCTGCAGCAAAAC 1105
      :::|||::::::       :::   :::   |||   |||:::|||
 276 yGlnLeuGluTyrGlnIleSerAlaAlaThrGlnThrAlaGlyAlaAlaT 293

1106 GCAGGCTGCCTCAGGGAAAGGAGCCTGGGTTGATTAACTTGTGTGCCAAT 1155
                |||
 293 hrSerSerAlaAlaGly.............................  298

1156 GTCCCACCCGTCCCAGGTAACATTTTGCCCCCTGAGGTCCGGGGTAATTT 1205
                                                 |||
 299 .....................................GlyAlaAl 301

1206 AATGGCTGCTGGA......CAAAACCTCCAAACTTCTGAAAGATCAGAAA 1249
       |||:::|||      |||:::|||   :::           :::::::
 301 aThrAlaGlyGlyGlySerGlnGlnLeuValGlnLeuProAlaGlySerI 318

1250 TGATAGCTACCTGGAGTCCAGCTGTACGGACACTGAGGAATATTACTAAT 1299
     :::::::|||   :::          |||              :::|||:::
 318 leValSerThrThrThrHis......ArgLeuSerValProValThrGlu 332

1300 AATGCTGACATTCAGCAGATCAACCGGCCATCAAATGTAGCACATATCTT 1349
      |||   :::|||   ::::::|||||      :::::::::   ::
 333 AlaAlaLeuValGlnLeuGlnArgArgProValTyrIleSerAsnAlaVa 349

1350 ACAGACTCTTTCAGCA.............................. 1365
     :      |||:::|||
 349 lGlyGlyLeuThrAlaGlyAlaThrTyrValArgMetProAlaAlaAlaV 366

1366 ..................CCTACGAAAAATTTAGAACAGCAG...... 1389
                        :::   |||:::   ||||||
 366 aIleSerArgSerProMetThrValLeuAsnValValGlnGlnAlaLeu 382

1390 ..............GTGAATCACAGCCAGCAGGGACATACAAATGCCAA 1424
                    :::   :::::|||||||   :::|||   |||
 383 ProAlaThrGlnLeuIleLeuGlnThrGlnGlnGlnGlnThrProAlaAl 399

1425 TGCAGTGCTGTTTAGCCAAGTGAAAGTCACTCCAGAGACACACATGCTAC 1474
      |||:::      :::|||:::   ::::::   |||   :::      |
 399 aAlaAlaAlaGluGlnGlnLeuAlaLeuAlaLeuGluGlnGlnArgGlnG 416

1475 AGCAGCAGCAGCAGGCCCAGCAGCAGCAGCAGCAGCACCCGGTTTTACAC 1524
     |||||:::|||||||||||||||||||:::||||||:::          :::
 416 lnGlnGluGlnGlnAlaGlnGlnGlnArgGlnGlnGlnGlnGlnGlnGlu 432

1525 CTTCAGCCCCAGCAGATAATGCAGCTCCAGCAGCAGCAGCAGCAGCAGAT 1574
       |||   ||||||   :::|||   ||||||||||||||||||||
 433 GlnGlnArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnAl 449

1575 CTCTCAGCAACCTTACCCCCAGCAGCCGCCGCATCCATTTTCACAGCAGC 1624
      :::|||:::         ||||||                 :::|||   |
 449 aGlnGlnGluGlnGlnArgGlnGlnGln............GlnGlnLeuG 462

1625 AGCAGCAGCAACAGCAGCCACCACCATCGCCTCAGCAGCATCAGCTTTTT 1674
     ||||||||||||||||          |||   :::|||:::
 462 lnGlnGlnGlnGlnGlnGlnGlnThrHisProValLysHisSerAlaSer 478
```

```
1675 GGA 1677
        |||
 479 Gly 479
```

seq_name: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG27784

seq_documentation_block:

ID ABG27784 standard; Protein; 90 AA.
XX
AC ABG27784;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #27775.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS91971.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 20; SEQ ID No 58143; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 90 AA;

alignment_scores:

Quality:	215.50	Length:	107
Ratio:	5.388	Gaps:	4
Percent Similarity:	37.383	Percent Identity:	33.645

```
alignment_block:
 US-09-664-641-10/rev x ABG27784   ..

 Align seg 1/1  to: ABG27784  from: 1  to: 90

    1655 GGCGATGGTGGTGGCTGCTGTTGCTGCTGCTGCTGCTGTGAAAATGGATG 1606
         |||   ||||||||||||||||||||||||||||||||||
       7 GlyGlyGlyGlyGlyCysCysCysCysCysCysCys..... 19

    1605 CGGCGGCTGCTGGGGGTAAGGTTGCTGAGAGATCTGCTGCTGCTGCTGCT 1556
                                          ||||||||||||||||
      20 ..............................CysCysCysCysCysC 25

    1555 GCTGGAGCTGCATTATCTGCTGGGGCTGAAGGTGTAAAACCGGGTGCTGC 1506
         ||
      25 ys.....................................||||||CysCys 27

    1505 TGCTGCTGCTGCTGGGCCTGCTGCTGCTGCTGTAGCATGTGTGTCTCTGG 1456
         ||||||||||||   :::||||||||||||||||
      28 CysCysCysCysCysCysCysCysCysCys.................. 38

    1455 AGTCACTTTCACTTGGCTAAACAGCACTGCATTGGCATTTGTATGTCCCT 1406

      38 ............................. . ................. 38

    1405 GCTGGCTGTGATTCACCTGCTGT.................TCTAAATTT 1374
                         ||||||                 ||||||
      39 ................CysCysTrpLysAspLeuArgAspSerLysAla 49

    1373 TTCGTAGGTGCTGAAAGAGTC 1353
         |||::::::   :::||||||
      50 PheIleSerPheSerArgVal 56
```

```
OM of: US-09-664-641-10 to: Pending_Patents_AA_New:*   out_format : pfs

Date: Jun 4, 2002  5:45 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL-frame+_n2p.model -DEV-xlp
 -Q-/cgn2_1/USPTO_spool/US09664641/runat_04062002_110933_9224/app_query.fasta_1.3697
 -DB-Pending_Patents_AA_New -QFMT-fastan -SUFFIX-rapn
 -GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
 -LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000
 -XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
 -YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
 -MATRIX-blosum62 -TRANS-human40.cdi -LIST-45 -DOCALIGN-200
 -THR_SCORE-pct -THR_MAX-100 -THR_MIN-0 -ALIGN-15 -MODE-LOCAL
 -OUTFMT-pfs -NORM-ext -HEAPSIZE-500 -MINLEN-0 -MAXLEN-2000000000
 -USER-US09664641_@CGN1_1_83 -NCPU-6 -JCPU-3 -LONGLOG
 -DEV_TIMEOUT-120 -WARN_TIMEOUT-30 -NO_XLPXY -WAIT -THREADS-1

Search information block:
Query: US-09-664-641-10
Query length: 3580
Database: Pending_Patents_AA_New:*
Database sequences: 160696
Database length: 45962370
Search time (sec): 96.670000



score_list:
Sequence                                          STrd Orig    ZScore    EScore Len  ! Documentation  ..
/cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep:US-09-300-008B-54 +  428.00  471.23  2.0e-18   8
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-123-155-381 -  275.00  276.58  5.3e-09   2
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-137-871-381 -  275.00  276.58  5.3e-09   2
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-141-761-381 -  275.00  276.58  5.3e-09   2
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-140-864-381 -  275.00  276.58  5.3e-09   2
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-140-923-381 -  275.00  276.58  5.3e-09   2
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-029-217A-24 -  272.50  268.90  8.1e-09   3
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-123-155-89 -   269.00  271.58  1.2e-08  18
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-063-545-149 -  269.00  271.58  1.2e-08   1
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-063-685-149 -  269.00  271.58  1.2e-08   1
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-137-871-89 -   269.00  271.58  1.2e-08  18
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-141-761-89 -   269.00  271.58  1.2e-08  18
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-140-864-89 -   269.00  271.58  1.2e-08  18
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-140-923-89 -   269.00  271.58  1.2e-08  18
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-123-155-147 -  266.50  262.82  1.8e-08   3
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-137-871-147 -  266.50  262.82  1.8e-08   3
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-141-761-147 -  266.50  262.82  1.8e-08   3
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-140-864-147 -  266.50  262.82  1.8e-08   3
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-140-923-147 -  266.50  262.82  1.8e-08   3
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-123-155-371 -  261.50  259.83  3.6e-08   2
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-137-871-371 -  261.50  259.83  3.6e-08   2
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-141-761-371 -  261.50  259.83  3.6e-08   2
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-140-864-371 -  261.50  259.83  3.6e-08   2
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-140-923-371 -  261.50  259.83  3.6e-08   2
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-123-155-99 -   260.00  261.80  4.1e-08  19
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-137-871-99 -   260.00  261.80  4.1e-08  19
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-141-761-99 -   260.00  261.80  4.1e-08  19
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-140-864-99 -   260.00  261.80  4.1e-08  19
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-140-923-99 -   260.00  261.80  4.1e-08  19
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-063-545-15 -   259.00  257.38  5.0e-08  27
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-063-685-15 -   259.00  257.38  5.0e-08  27
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-123-155-127 -  257.50  257.96  6.0e-08   2
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-137-871-127 -  257.50  257.96  6.0e-08   2
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-141-761-127 -  257.50  257.96  6.0e-08   2
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-140-864-127 -  257.50  257.96  6.0e-08   2
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-140-923-127 -  257.50  257.96  6.0e-08   2
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-063-545-53 -   257.00  259.69  6.2e-08  16
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-063-685-53 -   257.00  259.69  6.2e-08  16
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-123-155-543 -  257.00  252.58  6.9e-08   3
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-137-871-543 -  257.00  252.58  6.9e-08   3
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-141-761-543 -  257.00  252.58  6.9e-08
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-140-864-543 -  257.00  252.58  6.9e-08
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-140-923-543 -  257.00  252.58  6.9e-08
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-121-062-581 -  256.50  260.54  6.4e-08
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-123-155-291 -  255.50  254.70  8.0e-08

seq_name: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep:US-09-300-008B-54

seq_documentation_block:
; Sequence 54, Application US/09300008B
; GENERAL INFORMATION:
;  APPLICANT: Concannon et al.
;  TITLE OF INVENTION: A GENE ASSOCIATED WITH NIJMEGEN BREAKAGE
;  TITLE OF INVENTION:  SYNDROME, ITS GENE PRODUCT AND METHODS FOR THEIR USE
;  FILE REFERENCE: 9924-0003-228
;  CURRENT APPLICATION NUMBER: US/09/300,008B
;  CURRENT FILING DATE:  1999-04-27
;  PRIOR APPLICATION NUMBER: US 60/083,269
;  PRIOR FILING DATE: 1998-04-27
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 54
;   LENGTH: 84
;   TYPE: PRT
;   ORGANISM: Homo sapien
US-09-300-008B-54


alignment_scores:
           Quality:  428.00              Length:     84
             Ratio:  5.095                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 97.619

alignment_block:
 US-09-664-641-10 x US-09-300-008B-54   ..

 Align seq 1/1  to: US-09-300-008B-54  from: 1  to: 84

    2488 GAATTGACCCCTTTTGTGCTTTTCACTGGATTCGAGCCTGTCCAGGTTCA 2537
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 GluLeuThrProPheValLeuPheThrGlyPheGluProValGlnValGl 17

    2538 ACAGTATATTAAGAAGCTCTACATTCTTGGTGGAGAGGTTGCGGAGTCTG 2587
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      17 nGlnTyrIleLysLysLeuTyrIleLeuGlyGlyGluValAlaGluSerA 34

    2588 CACAGAAGTGCACACACCTCATTGCCAGCAAAGTGACTCGCACCGTGAAG 2637
         ||||||||||||||||||||||||||||||||||||||||||||:::|||
      34 laGlnLysCysThrHisLeuIleAlaSerLysValThrArgThrLeuLys 50

    2638 TTCCTGACGGCGATTTCTGTCGTGAAGCACATAGTGACGCCAGAGTGGCT 2687
         ||||||:::|||||||||||||||||||||||||||||||||||||||||
      51 PheLeuAlaAlaIleSerValValLysHisIleValThrProGluTrpLe 67

    2688 GGAAGAATGCTTCAGGTGTCAGAAGTTCATTGATGAGCAGAACTACATTC 2737
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 uGluGluCysPheArgCysGlnLysPheIleAspGluGlnAsnTyrIleL 84

    2738 TC 2739
         ||
      84 eu 84

seq_name: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-123-155-381

seq_documentation_block:
; Sequence 381, Application US/10123155
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gao,Wei-Qiang
```

```
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C30
;  CURRENT APPLICATION NUMBER: US/10/123,155
;  CURRENT FILING DATE:  2002-04-15
; Prior Application removed -  See Palm or File Wrapper
;  NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 381
;   LENGTH: 2236
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-10-123-155-381


alignment_scores:
            Quality:  275.00              Length:    131
              Ratio:  4.231                 Gaps:      4
 Percent Similarity:  49.618   Percent Identity: 40.458

alignment_block:
 US-09-664-641-10/rev x US-10-123-155-381   ..

 Align seq 1/1  to: US-10-123-155-381  from: 1  to: 2236

    1700 GGAATCTCCACTGCTGGATCATGTCCAAAAAGCTGATGCTGCTGAGGCGA 1651
         |||      |||:::::::::|||                 ||||||   |||
      20 GlyGlyCysThrGlyAlaGlyCysGlyCysCysThrCysCysThrGlyCy 36

    1650 TGGTGGTGGCTGCTGT......TGCTGC.................TGCT 1625
         ||||||   ||||||       ||||||                  ||||
      36 sGlyGlyCysCysCysGlyGlyCysCysThrGlyCysGlyCysGlyCysC 53

    1624 GCTGCTGTGAAAATGGATGCGGCGGCTGCTGGGGGTAAGGTTGCTGAGAG 1575
         ||||||||   :::   |||   ||||||   |||   ||||||
      53 ysCysCysGlyGlyCysCysCysGlyCysCysGlyCysGlyCysCysGly 69

    1574 ATCTGCTGC.........TGCTGCTGCTGCTGGAGCTGCATTATCTGCTG 1534
         |||||||          ||||||||||||   :::|||    |||
      70 CysCysCysAlaCysGlyCysCysCysCysAlaAlaCysCysCysCysGl 86

    1533 GGGCTGAAGGTGTAAAACCGGGTGCTGCTGCTGCTGCTGCTGGGCCTGCT 1484
         |||      |||      ||||||||||||||||||      :::||||
      86 yGlyCysCysCysGlyCysGlyCysCysCysCysThrAlaGlyCysC 103

    1483 GCTGCTGCTGTAGCATGTGTGTCTCTGGAGTCACTTTCACTTGGCTAAAC 1434
         ||||||||||:::   |||   :::|||
     103 ysCysCysCysGlyCysCysCysGlyGlyGlyCysCysCys..... 116

    1433 AGCACTGCATTGGCATTTGTATGTCCCTGCTGGCTGTGATTCACCTGCTG 1384
         :::   :::   :::      |||   |||                ||||
     117 GlyCysGlyCysCysCysGlyCysGlyCysCysCysGlyCysGlyCysCy 133

    1383 TTCTAAATTTTTCGTAGGTGCTGAAAGAGTCTGTAAGATATGT 1341
         |           :::||||||         |||      |||
     133 sCysAlaGlyGlyThrGlyAlaGlyCysGlyCysThrCysCys 147

seq_name: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-137-871-381

seq_documentation_block:
; Sequence 381, Application US/10137871
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gao,Wei-Qiang
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C153
;  CURRENT APPLICATION NUMBER: US/10/137,871
;  CURRENT FILING DATE:  2002-05-03
;  Prior Application removed - See Palm or File Wrapper
;  NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 381
;   LENGTH: 2236
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-10-137-871-381


alignment_scores:
            Quality:  275.00              Length:    131
              Ratio:  4.231                 Gaps:      4
 Percent Similarity:  49.618   Percent Identity: 40.458

alignment_block:
 US-09-664-641-10/rev x US-10-137-871-381   ..

 Align seq 1/1  to: US-10-137-871-381  from: 1  to: 2236

    1700 GGAATCTCCACTGCTGGATCATGTCCAAAAAGCTGATGCTGCTGAGGCGA 1651
         |||      |||:::::::::|||                 ||||||   |||
      20 GlyGlyCysThrGlyAlaGlyCysGlyCysCysThrCysCysThrGlyCy 36

    1650 TGGTGGTGGCTGCTGT......TGCTGC.................TGCT 1625
         ||||||   ||||||       ||||||                  ||||
      36 sGlyGlyCysCysCysGlyGlyCysCysThrGlyCysGlyCysGlyCysC 53

    1624 GCTGCTGTGAAAATGGATGCGGCGGCTGCTGGGGGTAAGGTTGCTGAGAG 1575
         ||||||||   :::   |||   ||||||   |||   ||||||
      53 ysCysCysGlyGlyCysCysCysGlyCysCysGlyCysGlyCysCysGly 69

    1574 ATCTGCTGC.........TGCTGCTGCTGCTGGAGCTGCATTATCTGCTG 1534
         ||||||           ||||||||||||   :::|||    |||
      70 CysCysCysAlaCysGlyCysCysCysCysAlaAlaCysCysCysCysGl 86

    1533 GGGCTGAAGGTGTAAAACCGGGTGCTGCTGCTGCTGCTGCTGGGCCTGCT 1484
         |||      |||      ||||||||||||||||||      :::||||
      86 yGlyCysCysCysGlyCysGlyCysCysCysCysThrAlaGlyCysC 103

    1483 GCTGCTGCTGTAGCATGTGTGTCTCTGGAGTCACTTTCACTTGGCTAAAC 1434
         ||||||||||:::   |||   :::|||
     103 ysCysCysCysGlyCysCysCysGlyGlyGlyCysCysCys..... 116

    1433 AGCACTGCATTGGCATTTGTATGTCCCTGCTGGCTGTGATTCACCTGCTG 1384
         :::   :::   :::      |||   |||                ||||
     117 GlyCysGlyCysCysCysGlyCysGlyCysCysCysGlyCysGlyCysCy 133

    1383 TTCTAAATTTTTCGTAGGTGCTGAAAGAGTCTGTAAGATATGT 1341
         |           :::||||||         |||      |||
     133 sCysAlaGlyGlyThrGlyAlaGlyCysGlyCysThrCysCys 147

seq_name: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-141-761-381
```

```
seq_documentation_block:
; Sequence 381, Application US/10141761
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gao,Wei-Qiang
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C198
;  CURRENT APPLICATION NUMBER: US/10/141,761
;  CURRENT FILING DATE:  2002-05-08
;  Prior Application removed - See Palm or File Wrapper
;  NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 381
;   LENGTH: 2236
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-10-141-761-381


alignment_scores:
           Quality:  275.00              Length:    131
             Ratio:  4.231                Gaps:      4
 Percent Similarity:  49.618   Percent Identity: 40.458

alignment_block:
 US-09-664-641-10/rev x US-10-141-761-381   ..

 Align seg 1/1  to: US-10-141-761-381  from: 1  to: 2236

    1700 GGAATCTCCACTGCTGGATCATGTCCAAAAAGCTGATGCTGCTGAGGCGA 1651
         |||      |||:::::::::|||          ||||||   |||
      20 GlyGlyCysThrGlyAlaGlyCysGlyCysCysThrCysCysThrGlyCy 36

    1650 TGGTGGTGGCTGCTGT......TGCTGC.................TGCT 1625
         ||||||  ||||||      ||||||                   ||||
      36 sGlyGlyCysCysCysGlyGlyCysCysThrGlyCysGlyCysGlyCysC 53

    1624 GCTGCTGTGAAAATGGATGCGGCGGCTGCTGGGGGTAAGGTTGCTGAGAG 1575
         ||||||||  :::  |||  ||||||  |||  ||||||
      53 ysCysCysGlyGlyCysCysCysGlyCysCysGlyCysGlyCysCysGly 69

    1574 ATCTGCTGC.........TGCTGCTGCTGCTGGAGCTGCATTATCTGCTG 1534
           ||||||          ||||||||||||   :::|||      |||
      70 CysCysCysAlaCysGlyCysCysCysCysAlaAlaCysCysCysCysGl 86

    1533 GGGCTGAAGGTGTAAAACCGGGTGCTGCTGCTGCTGCTGCTGGGCCTGCT 1484
         |||      |||      ||||||||||||||||||      :::||||
      86 yGlyCysCysCysGlyCysGlyCysCysCysCysCysThrAlaGlyCysC 103

    1483 GCTGCTGCTGTAGCATGTGTGTCTCTGGAGTCACTTTCACTTGGCTAAAC 1434
         ||||||||||:::   |||   :::|||
     103 ysCysCysCysGlyCysCysCysGlyGlyGlyCysCysCys......... 116

    1433 AGCACTGCATTGGCATTTGTATGTCCCTGCTGGCTGTGATTCACCTGCTG 1384
         :::   :::   :::     |||  |||               |||||
     117 GlyCysGlyCysCysCysGlyCysGlyCysCysCysGlyCysGlyCysCy 133

    1383 TTCTAAATTTTTCGTAGGTGCTGAAAGAGTCTGTAACATATGT 1341
         |          :::|||||||         |||      |||
     133 sCysAlaGlyGlyThrGlyAlaGlyCysGlyCysThrCysCys 147

seq_name: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-140-864-381

seq_documentation_block:
; Sequence 381, Application US/10140864
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gao,Wei-Qiang
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C184
;  CURRENT APPLICATION NUMBER: US/10/140,864
;  CURRENT FILING DATE:  2002-05-07
;  Prior Application removed - See Palm or File Wrapper
;  NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 381
;   LENGTH: 2236
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-10-140-864-381


alignment_scores:
           Quality:  275.00              Length:    131
             Ratio:  4.231                Gaps:      4
 Percent Similarity:  49.618   Percent Identity: 40.458

alignment_block:
 US-09-664-641-10/rev x US-10-140-864-381   ..

 Align seg 1/1  to: US-10-140-864-381  from: 1  to: 2236

    1700 GGAATCTCCACTGCTGGATCATGTCCAAAAAGCTGATGCTGCTGAGGCGA 1651
         |||      |||:::::::::|||          ||||||   |||
      20 GlyGlyCysThrGlyAlaGlyCysGlyCysCysThrCysCysThrGlyCy 36

    1650 TGGTGGTGGCTGCTGT......TGCTGC.................TGCT 1625
         ||||||  ||||||      ||||||                   ||||
      36 sGlyGlyCysCysCysGlyGlyCysCysThrGlyCysGlyCysGlyCysC 53

    1624 GCTGCTGTGAAAATGGATGCGGCGGCTGCTGGGGGTAAGGTTGCTGAGAG 1575
         ||||||||  :::  |||  ||||||  |||  ||||||
      53 ysCysCysGlyGlyCysCysCysGlyCysCysGlyCysGlyCysCysGly 69

    1574 ATCTGCTGC.........TGCTGCTGCTGCTGGAGCTGCATTATCTGCTG 1534
           ||||||          ||||||||||||   :::|||      |||
      70 CysCysCysAlaCysGlyCysCysCysCysAlaAlaCysCysCysCysGl 86

    1533 GGGCTGAAGGTGTAAAACCGGGTGCTGCTGCTGCTGCTGCTGGGCCTGCT 1484
         |||      |||      ||||||||||||||||||      :::||||
      86 yGlyCysCysCysGlyCysGlyCysCysCysCysCysThrAlaGlyCysC 103

    1483 GCTGCTGCTGTAGCATGTGTGTCTCTGGAGTCACTTTCACTTGGCTAAAC 1434
         ||||||||||:::   |||   :::|||
     103 ysCysCysCysGlyCysCysCysGlyGlyGlyCysCysCys......... 116
```

```
 1433 AGCACTGCATTGGCATTTGTATGTCCCTGCTGGCTGTGATTCACCTGCTG 1384
      :::  :::  :::     |||  |||              |||||
  117 GlyCysGlyCysCysCysGlyCysGlyCysCysCysGlyCysGlyCysCy 133

 1383 TTCTAAATTTTTCGTAGGTGCTGAAAGAGTCTGTAAGATATGT 1341
      |          :::||||||       |||     |||
  133 sCysAlaGlyGlyThrGlyAlaGlyCysGlyCysThrCysCys 147

seq_name: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-140-923-381

seq_documentation_block:
; Sequence 381, Application US/10140923
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gao,Wei-Qiang
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C188
;  CURRENT APPLICATION NUMBER: US/10/140,923
;  CURRENT FILING DATE:  2002-05-07
;  Prior Application removed - See Palm or File Wrapper
;  NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 381
;   LENGTH: 2236
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-10-140-923-381


alignment_scores:
            Quality:  275.00          Length:    131
              Ratio:  4.231             Gaps:      4
 Percent Similarity:  49.618   Percent Identity: 40.458

alignment_block:
 US-09-664-641-10/rev x US-10-140-923-381   ..

 Align seq 1/1  to: US-10-140-923-381  from: 1  to: 2236

 1700 GGAATCTCCACTGCTGGATCATGTCCAAAAAGCTGATGCTGCTGAGGCGA 1651
      |||     |||||||||||||          ||||||  |||
   20 GlyGlyCysThrGlyAlaGlyCysGlyCysCysThrCysCysThrGlyCy 36

 1650 TGGTGGTGGCTGCTGT......TGCTGC.................TGCT 1625
      ||||||  ||||||    ||||||                    ||||
   36 sGlyGlyCysCysCysGlyGlyCysCysThrGlyCysGlyCysGlyCysC 53

 1624 GCTGCTGTGAAAATGGATGCCGGCGGCTGCTGGGGGTAAGGTTGCTGAGAG 1575
      ||||||||  :::  |||  ||||||  |||  ||||||
   53 ysCysCysGlyGlyCysCysCysGlyCysCysGlyCysGlyCysCysGly 69

 1574 ATCTGCTGC.....TGCTGCTGCTGCTGGAGCTGCATTATCTGCTG 1534
        ||||||       ||||||||||||  :::|||     |||
   70 CysCysCysAlaCysGlyCysCysCysCysAlaAlaCysCysCysGl 86

 1533 GGGCTGAAGGTGTAAAACCGGGTGCTGCTGCTGCTGCTGGGCCTGCT 1484
      |||     |||     ||||||||||||||||||     :::||||
```

```
   86 yGlyCysCysCysGlyCysGlyCysCysCysCysThrAlaGlyCysC 103

 1483 GCTGCTGCTGTAGCATGTGTGTCTCTGGAGTCACTTTCACTTGGCTAAAC 1434
      ||||||||||:::   |||  :::|||
  103 ysCysCysCysGlyCysCysCysGlyGlyGlyCysCysCys..... 116

 1433 AGCACTGCATTGGCATTTGTATGTCCCTGCTGGCTGTGATTCACCTGCTG 1384
      :::  :::  :::     |||  |||              |||||
  117 GlyCysGlyCysCysCysGlyCysGlyCysCysCysGlyCysGlyCysCy 133

 1383 TTCTAAATTTTTCGTAGGTGCTGAAAGAGTCTGTAAGATATGT 1341
      |          :::||||||       |||     |||
  133 sCysAlaGlyGlyThrGlyAlaGlyCysGlyCysThrCysCys 147

seq_name: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-029-217A-24

seq_documentation_block:
; Sequence 24, Application US/10029217A
; GENERAL INFORMATION:
;  APPLICANT: OLSON, ERIC N.
;  APPLICANT:  WANG, DA-ZHI
;  TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING TO A CARDIAC-SPECIFIC
;  TITLE OF INVENTION:  NUCLEAR REGULATORY FACTOR
;  FILE REFERENCE: UTSD:695US
;  CURRENT APPLICATION NUMBER: US/10/029,217A
;  CURRENT FILING DATE:  2002-03-19
;  PRIOR APPLICATION NUMBER: 60/257,761
;  PRIOR FILING DATE: 2000-12-21
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
;   LENGTH: 3907
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-10-029-217A-24


alignment_scores:
            Quality:  272.50          Length:    169
              Ratio:  3.406             Gaps:      7
 Percent Similarity:  47.337   Percent Identity: 34.320

alignment_block:
 US-09-664-641-10/rev x US-10-029-217A-24   ..

 Align seq 1/1  to: US-10-029-217A-24  from: 1  to: 3907

 1778 GCCAGCAGTTGCTTATCAGACATCTGCTCTGGATAATCCGCAATTGCAAA 1729
      |||:::  |||  :::     |||:::|||  :::|||
 2317 AlaThrCysCysCysThrGlyCysCysAlaGlyGlyGlyAla........ 2330

 1728 CACACATCCCAATAAGAAGCCTTCTTCTGGAATCTCCACTGCTGGATCAT 1679
                          ::::::|||  ::::::||||||  |
 2331 .....................AlaGlyGlyAlaGlyAlaAlaGlyCysC 2340

 1678 GTCCAAAAAGCTGATGCTGCTGAGGCGATGGT..................... 1647
      ||          ||||||  :::  |||
 2340 ysAlaThrCysCysCysCysGlyAlaAlaGlyAlaCysAlaGlyThrCys 2356

 1646 .........GGTGGCTGCTGTTGCTGCTGCTGCTGCTGTGAAAATGGATG 1606
              ||||||      ||||||||||||||||||  :::||||
 2357 ThrGlyThrGlyGlyGlyThrCysCysCysCysCysCysThrGlyGlyCy 2373

 1605 CGGCGGCTGCTGGGGGTAAGGTTGCTGAGAGATCTGCTGCTGCTGCTGCT 1556
      |:::||||||         ||||||          :::  |||:::||||
 2373 sAlaGlyCysAlaCysAlaGlyCys......CysAlaThrCysAlaCysC 2388

 1555 GCTGGAGCTGCATTATCTGCTGGGGCTGAAGGTGTAAAACCGGGTGCTGC 1506
      ||   :::|||      |||  |||      |||         ||||||
 2388 ysThrThrCysThrGlyCysThrGlyAlaGlyCysThrCysCysCysCys 2404

 1505 TGC.................TGCTGCTGCTGGGCCTGCTGCTGCTGCTG 1474
      |||                 |||||||||  |||||||||  |||||
```

```
    2405 CysAlaGlyGlyCysThrGlyCysCysCysCysAlaCysCysThrCysCy 2421

    1473 TAGCATGTGTGTCTCTGGAGTCACTTTCACTTGGCTAAACAGCACTGCAT 1424
         |:::   |||::::::|||   |||                     |||
    2421 sThrCysCysAlaGlyGlyCysThr.................CysAlaC 2432

    1423 TGGCATTTGTATGTCCCTGCTGGCTGTGATTCACCTGCTGTTCTAAATTT 1374
           :::   :::|||   |||           |||||||||
    2432 ysCysCysThrCysCysCys............ThrCysCysCysThrGly 2444

    1373 TTCCTA.................GCTGCTGAAAGAGTCTGTAAGATATG 1342
           :::                |||||||      :::|||      ||
    2445 GlyAlaCysGlyCysCysThrGlyGlyAlaGlyGlyAlaCysThrThrCy 2461

    1341 TGCTACA 1335
         |:::|||
    2461 sCysThr 2463

seq_name: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-123-155-89

seq_documentation_block:
; Sequence 89, Application US/10123155
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gao,Wei-Qiang
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C30
;  CURRENT APPLICATION NUMBER: US/10/123,155
;  CURRENT FILING DATE:  2002-04-15
; Prior Application removed -  See Palm or File Wrapper
;  NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 89
;   LENGTH: 1893
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-10-123-155-89


alignment_scores:
            Quality:  269.00            Length:    168
              Ratio:  3.494               Gaps:      8
 Percent Similarity:  45.833   Percent Identity: 33.333

alignment_block:
 US-09-664-641-10/rev x US-10-123-155-89   ..

 Align seg 1/1  to: US-10-123-155-89  from: 1  to: 1893

    1703 TCTGGAATCTCCACTGCTGGATCATGTCCAAAAAGCTGATGCTGCTGAGG 1654
         :::|||      |||:::|||:::|||           ||||||   ||
      60 AlaGlyAlaCysThrThrGlyAlaCysThr.........CysCysCysGl 73

    1653 CGATGGTGGTGGCTGCTGTTGCTGC..................... 1629
         |   |||   ||||||||||||||||
      73 yCysGlyCysGlyCysCysCysCysAlaAlaCysCysCysThrGlyCysT 90

    1628 ..........TGCTGCTGC..........................  1620
                  |||||||||
      90 hrThrAlaThrCysCysCysThrThrGlyAlaCysCysGlyThrCysGly 106

    1619 ..................................TGTGAAAATGGATG 1606
                                           |||   :::|||||
     107 AlaGlyThrGlyThrCysAlaGlyAlaGlyAlaThrCysCysThrGlyCy 123

    1605 CGGCGGCTGCTGGGGGTAAGGTTGCTGAGAGATCTGCTGCTGC......T 1562
         |:::||||||   |||      |||         |||||||||      |
     123 sAlaGlyCysCysGlyCysCysCysAlaGlyThrCysCysCysGlyGlyC 140

    1561 GCTGCTGCTGGAGCTGCATTATCTGCTGGGGCTGAAGGTGTAAAACCGGG 1512
         ||||||||   :::|||      |||                    |||
     140 ysCysCysCysThrCysThrCysCys.................CysGly 150

    1511 TGCTGCTGCTGCTGCTGCTGGGCCTGCTGC...TGCTGCTGTAGCATGTG 1465
         ||||||||||||::::|||   :::||||||   |||||||||:::   ||
     151 CysCysCysCysAlaCysAlaCysCysCysAlaCysCysCysThrCysCy 167

    1464 TGTCTCTGGAGTCACT...TTCACTTGGCTAAACAGCACTGCATTGGCAT 1418
         |:::::|||   |||      |||      ::::::|||:::   :::
     167 sThrGlyGlyCysThrCysThrThrCysCysThrGlyThrThrThrThrT 184

    1417 TTGTATGTCCCTGCTGGCTGTGATTCACCTGCTGTTCTAAATTTTTCGTA 1368
           :::|||   |||           |||||||||:::
     184 hrAlaCysThrCys.........CysThrCysCysThrThrThrThrCys 197

    1367 GGTGCTGAAAGAGTCTGTAAGATATGTGCTACATTTGATGGCCGGTTCAT 1318
         ::::::      :::         ||||||:::      |||
     198 AlaThrThrCysAlaThrAlaAlaCysAlaAlaAlaAlaGlyCysThrAl 214

    1317 CTGC 1314
          |||
     214 aCys 215

seq_name: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-063-545-149

seq_documentation_block:
; Sequence 149, Application US/10063545
; GENERAL INFORMATION:
;  APPLICANT: Eaton,Dan L.
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Grimaldi,Christopher J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3230R1C1
;  CURRENT APPLICATION NUMBER: US/10/063,545
;  CURRENT FILING DATE:  2002-05-02
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 149
;   LENGTH: 1893
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-10-063-545-149


alignment_scores:
            Quality:  269.00            Length:    168
              Ratio:  3.494               Gaps:      8
 Percent Similarity:  45.833   Percent Identity: 33.333

alignment_block:
 US-09-664-641-10/rev x US-10-063-545-149   ..

 Align seg 1/1  to: US-10-063-545-149  from: 1  to: 1893
```

```
    1703 TCTGGAATCTCCACTGCTGGATCATGTCCAAAAAGCTGATGCTGCTGAGG 1654
           :::|||      |||:::|||:::|||           ||||||   ||
      60 AlaGlyAlaCysThrThrGlyAlaCysThr.........CysCysCysGl 73

    1653 CGATGGTGGTGGCTGCTGTTGCTGC..................... 1629
           |   |||   ||||||||||||||||
      73 yCysGlyCysGlyCysCysCysCysAlaAlaCysCysCysThrGlyCysT 90

    1628 ..........TGCTGCTGC.......................... 1620
                   |||||||||
      90 hrThrAlaThrCysCysCysThrThrGlyAlaCysCysGlyThrCysGly 106

    1619 ................................TGTGAAAATGGATG 1606
                                         |||   :::|||||
     107 AlaGlyThrGlyThrCysAlaGlyAlaGlyAlaThrCysCysThrGlyCy 123

    1605 CGGCGGCTGCTGGGGGTAAGGTTGCTGAGAGATCTGCTGCTGC......T 1562
           |:::||||||   |||      |||         |||||||||      |
     123 sAlaGlyCysCysGlyCysCysCysAlaGlyThrCysCysCysGlyGlyC 140

    1561 GCTGCTGCTGGAGCTGCATTATCTGCTGGG::TGAAGGTGTAAAACCGGG 1512
           ||||||||   :::|||      |||                   |||
     140 ysCysCysCysThrCysThrCysCys.....................CysGly 150

    1511 TGCTGCTGCTGCTGCTGCTGGGCCTGCTGC.. TGCTGCTGTAGCATGTG 1465
           ||||||||||||:::|||   :::||||||   |||||||||:::   ||
     151 CysCysCysCysAlaCysAlaCysCysCysAlaCysCysCysThrCysCy 167

    1464 TGTCTCTGGAGTCACT...TTCACTTGGCTAAACAGCACTGCATTGGCAT 1418
           |:::::|||   |||      |||      ::::::|||:::   :::
     167 sThrGlyGlyCysThrCysThrThrCysCysThrGlyThrThrThrThrT 184

    1417 TTGTATGTCCCTGCTGGCTGTGATTCACCTGCTGTTCTAAATTTTTCGTA 1368
             :::|||   |||           |||||||||:::
     184 hrAlaCysThrCys.........CysThrCysCysThrThrThrThrCys 197

    1367 GGTGCTGAAAGAGTCTGTAAGATATGTGCTACATTTGATGGCCGGTTCAT 1318
           ::::::      :::          ||||||:::      |||
     198 AlaThrThrCysAlaThrAlaAlaCysAlaAlaAlaAlaGlyCysThrAl 214

    1317 CTGC 1314
           |||
     214 aCys 215
```

seq_name: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-063-685-149

```
seq_documentation_block:
; Sequence 149, Application US/10063685
; GENERAL INFORMATION:
;  APPLICANT: Eaton,Dan L.
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Grimaldi,Christopher J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3230R1C1
;  CURRENT APPLICATION NUMBER: US/10/063,685
;  CURRENT FILING DATE:  2002-05-08
;      Prior Application removed - See Palm or File Wrapper
;  NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 149
;   LENGTH: 1893
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-10-063-685-149


alignment_scores:
```

```
           Quality:  269.00          Length:    168
             Ratio:  3.494             Gaps:      8
Percent Similarity:  45.833   Percent Identity: 33.333

alignment_block:
 US-09-664-641-10/rev x US-10-063-685-149   ..

 Align seg 1/1  to: US-10-063-685-149  from: 1  to: 1893

    1703 TCTGGAATCTCCACTGCTGGATCATGTCCAAAAAGCTGATGCTGCTGAGG 1654
           :::|||      |||:::|||:::|||           ||||||   ||
      60 AlaGlyAlaCysThrThrGlyAlaCysThr.........CysCysCysGl 73

    1653 CGATGGTGGTGGCTGCTGTTGCTGC..................... 1629
           |   |||   ||||||||||||||||
      73 yCysGlyCysGlyCysCysCysCysAlaAlaCysCysCysThrGlyCysT 90

    1628 ..........TGCTGCTGC.......................... 1620
                   |||||||||
      90 hrThrAlaThrCysCysCysThrThrGlyAlaCysCysGlyThrCysGly 106

    1619 ................................TGTGAAAATGGATG 1606
                                         |||   :::|||||
     107 AlaGlyThrGlyThrCysAlaGlyAlaGlyAlaThrCysCysThrGlyCy 123

    1605 CGGCGGCTGCTGGGGGTAAGGTTGCTGAGAGATCTGCTGCTGC......T 1562
           |:::||||||   |||      |||         |||||||||      |
     123 sAlaGlyCysCysGlyCysCysCysAlaGlyThrCysCysCysGlyGlyC 140

    1561 GCTGCTGCTGGAGCTGCATTATCTGCTGGGGCTGAAGGTGTAAAACCGGG 1512
           ||||||||   :::|||      |||                   |||
     140 ysCysCysCysThrCysThrCysCys.....................CysGly 150

    1511 TGCTGCTGCTGCTGCTGCTGGGCCTGCTGC...TGCTGCTGTAGCATGTG 1465
           ||||||||||||:::|||   :::||||||   |||||||||:::   ||
     151 CysCysCysCysAlaCysAlaCysCysCysAlaCysCysCysThrCysCy 167

    1464 TGTCTCTGGAGTCACT...TTCACTTGGCTAAACAGCACTGCATTGGCAT 1418
           |:::::|||   |||      |||      ::::::|||:::   :::
     167 sThrGlyGlyCysThrCysThrThrCysCysThrGlyThrThrThrThrT 184

    1417 TTGTATGTCCCTGCTGGCTGTGATTCACCTGCTGTTCTAAATTTTTCGTA 1368
             :::|||   |||           |||||||||:::
     184 hrAlaCysThrCys.........CysThrCysCysThrThrThrThrCys 197

    1367 GGTGCTGAAAGAGTCTGTAAGATATGTGCTACATTTGATGGCCGGTTCAT 1318
           ::::::      :::          ||||||:::      |||
     198 AlaThrThrCysAlaThrAlaAlaCysAlaAlaAlaAlaGlyCysThrAl 214

    1317 CTGC 1314
           |||
     214 aCys 215
```

seq_name: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-137-871-89

```
seq_documentation_block:
; Sequence 89, Application US/10137871
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gao,Wei-Qiang
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
```

```
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C153
;  CURRENT APPLICATION NUMBER: US/10/137,871
;  CURRENT FILING DATE:  2002-05-03
;  Prior Application removed - See Palm or File Wrapper
;  NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 89
;   LENGTH: 1893
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-10-137-871-89


alignment_scores:
            Quality:  269.00            Length:    168
              Ratio:  3.494               Gaps:      8
 Percent Similarity:  45.833   Percent Identity: 33.333

alignment_block:
 US-09-664-641-10/rev x US-10-137-871-89   ..

 Align seg 1/1  to: US-10-137-871-89  from: 1  to: 1893

    1703 TCTGGAATCTCCACTGCTGGATCATGTCCAAAAAGCTGATGCTGCTGAGG 1654
         ::::||      |||:::|||:::|||          ||||||   ||
      60 AlaGlyAlaCysThrThrGlyAlaCysThr.....CysCysCysGl 73

    1653 CGATGGTGGTGGCTGCTGTTGCTGC..................... 1629
         |   |||   ||||||||||||||||
      73 yCysGlyCysGlyCysCysCysCysAlaAlaCysCysCysThrGlyCysT 90

    1628 ..........TGCTGCTGC.......................... 1620
                   |||||||||
      90 hrThrAlaThrCysCysCysThrThrGlyAlaCysCysGlyThrCysGly 106

    1619 ................................TGTGAAAATGGATG 1606
                                         |||   :::||||||
     107 AlaGlyThrGlyThrCysAlaGlyAlaGlyAlaThrCysCysThrGlyCy 123

    1605 CGGCGGCTGCTGGGGGTAACGTTGCTGAGACATCTGCTGCTGC......T 1562
         |:::||||||   |||      |||         |||||||||      |
     123 sAlaGlyCysCysGlyCysCysCysAlaGlyThrCysCysCysGlyGlyC 140

    1561 GCTGCTGCTGGAGCTGCATTATCTGCTGGGGCTGAAGGTGTAAAACCGGG 1512
         ||||||||   :::|||      |||                  |||
     140 ysCysCysCysThrCysThrCysCys.................CysGly 150

    1511 TGCTGCTGCTGCTGCTGCTGGGCCTGCTGC...TGCTGCTGTAGCATGTG 1465
         ||||||||||||:::|||   :::|||||||   ||||||||:::   ||
     151 CysCysCysCysAlaCysAlaCysCysCysAlaCysCysCysThrCysCy 167

    1464 TGTCTCTGGAGTCACT...TTCACTTGGCTAAACAGCACTGCATTGGCAT 1418
         |:::::|||   |||      |||      ::::::|||:::   :::
     167 sThrGlyGlyCysThrCysThrThrCysCysThrGlyThrThrThrThrT 184

    1417 TTGTATGTCCCTGCTGGCTGTGATTCACCTGCTGTTCTAAATTTTTCGTA 1368
           :::|||   |||             ||||||||||:::
     184 hrAlaCysThrCys.........CysThrCysCysThrThrThrThrCys 197

    1367 GGTGCTGAAAGAGTCTGTAAGATATGTGCTACATTTGATGGCCGGTTCAT 1318
         ::::::      :::         |||||:::      |||
     198 AlaThrThrCysAlaThrAlaAlaCysAlaAlaAlaAlaGlyCysThrAl 214

    1317 CTGC 1314
          |||
     214 aCys 215

seq_name: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-141-761-89

seq_documentation_block:
```

```
; Sequence 89, Application US/10141761
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gao,Wei-Qiang
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C198
;  CURRENT APPLICATION NUMBER: US/10/141,761
;  CURRENT FILING DATE:  2002-05-08
;  Prior Application removed - See Palm or File Wrapper
;  NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 89
;   LENGTH: 1893
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-10-141-761-89


alignment_scores:
            Quality:  269.00            Length:    168
              Ratio:  3.494               Gaps:      8
 Percent Similarity:  45.833   Percent Identity: 33.333

alignment_block:
 US-09-664-641-10/rev x US-10-141-761-89   ..

 Align seg 1/1  to: US-10-141-761-89  from: 1  to: 1893

    1703 TCTGGAATCTCCACTGCTGGATCATGTCCAAAAAGCTGATGCTGCTGAGG 1654
         :::|||      |||:::|||:::|||          ||||||   ||
      60 AlaGlyAlaCysThrThrGlyAlaCysThr.....CysCysCysGl 73

    1653 CGATGGTGGTGGCTGCTGTTGCTGC..................... 1629
         |   |||   ||||||||||||||||
      73 yCysGlyCysGlyCysCysCysCysAlaAlaCysCysCysThrGlyCysT 90

    1628 ..........TGCTGCTGC.......................... 1620
                   |||||||||
      90 hrThrAlaThrCysCysCysThrThrGlyAlaCysCysGlyThrCysGly 106

    1619 ................................TGTGAAAATGGATG 1606
                                         |||   :::||||||
     107 AlaGlyThrGlyThrCysAlaGlyAlaGlyAlaThrCysCysThrGlyCy 123

    1605 CGGCGGCTGCTGGGGGTAAGGTTGCTGAGACATCTGCTGCTGC......T 1562
         |:::||||||   |||      |||         |||||||||      |
     123 sAlaGlyCysCysGlyCysCysCysAlaGlyThrCysCysCysGlyGlyC 140

    1561 GCTGCTGCTGGAGCTGCATTATCTGCTGGGGCTGAAGGTGTAAAACCGGG 1512
         ||||||||   :::|||      |||                  |||
     140 ysCysCysCysThrCysThrCysCys.................CysGly 150

    1511 TGCTGCTGCTGCTGCTGCTGGGCCTGCTGC...TGCTGCTGTAGCATGTG 1465
         ||||||||||||:::|||   :::|||||||   ||||||||:::   ||
     151 CysCysCysCysAlaCysAlaCysCysCysAlaCysCysCysThrCysCy 167

    1464 TGTCTCTGGAGTCACT...TTCACTTGGCTAAACAGCACTGCATTGGCAT 1418
         |:::::|||   |||      |||      ::::::|||:::   :::
```

```
     167 sThrGlyGlyCysThrCysThrThrCysCysThrGlyThrThrThrThrT 184

    1417 TTGTATGTCCCTGCTGGCTGTGATTCACCTGCTGTTCTAAATTTTTCGTA 1368
           :::|||   |||           |||||||||:::
     184 hrAlaCysThrCys.........CysThrCysCysThrThrThrThrCys 197

    1367 GGTGCTGAAAGAGTCTGTAAGATATGTGCTACATTTGATGGCCGGTTCAT 1318
         ::::::      :::         ||||||:::      |||
     198 AlaThrThrCysAlaThrAlaAlaCysAlaAlaAlaAlaGlyCysThrAl 214

    1317 CTGC 1314
          |||
     214 aCys 215

seq_name: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-140-864-89

seq_documentation_block:
; Sequence 89, Application US/10140864
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gao,Wei-Qiang
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C184
;  CURRENT APPLICATION NUMBER: US/10/140,864
;  CURRENT FILING DATE:  2002-05-07
;  Prior Application removed - See Palm or File Wrapper
;  NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 89
;   LENGTH: 1893
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-10-140-864-89


alignment_scores:
            Quality:  269.00            Length:    168
              Ratio:  3.494               Gaps:      8
 Percent Similarity:  45.833   Percent Identity: 33.333

alignment_block:
 US-09-664-641-10/rev x US-10-140-864-89   ..

 Align seg 1/1  to: US-10-140-864-89  from: 1  to: 1893

    1703 TCTGGAATCTCCACTGCTGGATCATGTCCAAAAAGCTGATGCTGCTGAGG 1654
         :::|||      |||:::|||:::|||           ||||||   ||
      60 AlaGlyAlaCysThrThrGlyAlaCysThr.........CysCysCysGl 73

    1653 CGATGGTGGTGGCTGCTGTTGCTGC..................... 1629
         !   |||   ||||||||||||||||
      73 yCysGlyCysGlyCysCysCysCysAlaAlaCysCysCysThrGlyCysT 90

    1628 ..........TCCTGCTGC..........................  1620
                   |||||||||
      90 hrThrAlaThrCysCysCysThrThrGlyAlaCysCysGlyThrCysGly 106

    1619 ................................TGTGAAAATGGATG 1606
```

```
                                                |||   :::|||||
     107 AlaGlyThrGlyThrCysAlaGlyAlaGlyAlaThrCysCysThrGlyCy 123

    1605 CGGCGGCTGCTGGGGGTAAGGTTGCTGAGAGATCTGCTGCTGC......T 1562
         |:::||||||   |||      |||         |||||||||      |
     123 sAlaGlyCysCysGlyCysCysCysAlaGlyThrCysCysCysGlyGlyC 140

    1561 GCTGCTGCTGGAGCTGCATTATCTGCTGGGGCTGAAGGTGTAAAACCGGG 1512
         ||||||||   :::|||      |||                    |||
     140 ysCysCysCysThrCysThrCysCys.................CysGly 150

    1511 TGCTGCTGCTGCTGCTGCTGGGCCTGCTGC...TGCTGCTGTAGCATGTG 1465
         ||||||||||||||:::|||   :::||||||   |||||||||:::   ||
     151 CysCysCysCysAlaCysAlaCysCysCysAlaCysCysCysThrCysCy 167

    1464 TGTCTCTGGAGTCACT...TTCACTTGGCTAAACAGCACTGCATTGGCAT 1418
         |:::::::|||   |||      |||      ::::::|||:::   :::
     167 sThrGlyGlyCysThrCysThrThrCysCysThrGlyThrThrThrThrT 184

    1417 TTGTATGTCCCTGCTGGCTGTGATTCACCTGCTGTTCTAAATTTTTCGTA 1368
           :::|||   |||           |||||||||:::
     184 hrAlaCysThrCys.........CysThrCysCysThrThrThrThrCys 197

    1367 GGTGCTGAAAGAGTCTGTAAGATATGTGCTACATTTGATGGCCGGTTCAT 1318
         ::::::      :::         ||||||:::      |||
     198 AlaThrThrCysAlaThrAlaAlaCysAlaAlaAlaAlaGlyCysThrAl 214

    1317 CTGC 1314
          |||
     214 aCys 215

seq_name: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-140-923-89

seq_documentation_block:
; Sequence 89, Application US/10140923
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gao,Wei-Qiang
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C188
;  CURRENT APPLICATION NUMBER: US/10/140,923
;  CURRENT FILING DATE:  2002-05-07
;  Prior Application removed - See Palm or File Wrapper
;  NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 89
;   LENGTH: 1893
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-10-140-923-89


alignment_scores:
            Quality:  269.00            Length:    168
              Ratio:  3.494               Gaps:      8
 Percent Similarity:  45.833   Percent Identity: 33.333

alignment_block:
```

```
US-09-664-641-10/rev x US-10-140-923-89  ..

Align seq 1/1  to: US-10-140-923-89  from: 1  to: 1893

   1703 TCTGGAATCTCCACTGCTGGATCATGTCCAAAAAGCTGATGCTGCTGAGG 1654
        :::|||      |||:::|||:::|||           ||||||   ||
     60 AlaGlyAlaCysThrThrGlyAlaCysThr.....CysCysCysGl 73

   1653 CGATGGTGGTGGCTGCTGTTGCTGC..................... 1629
        |   |||   ||||||||||||||||
     73 yCysGlyCysGlyCysCysCysCysAlaAlaCysCysCysThrGlyCysT 90

   1628 ..........TGCTGCTGC..........................  1620
                  |||||||||
     90 hrThrAlaThrCysCysCysThrThrGlyAlaCysCysGlyThrCysGly 106

   1619 ................................TGTGAAAATGGATG 1606
                                        |||   :::||||||
    107 AlaGlyThrGlyThrCysAlaGlyAlaGlyAlaThrCysCysThrGlyCy 123

   1605 CGGCGGCTGCTGGGGGTAAGGTTGCTGAGAGATCTGCTGCTGC......T 1562
        |:::|||||||   |||      |||         |||||||||      |
    123 sAlaGlyCysCysGlyCysCysCysAlaGlyThrCysCysCysGlyGlyC 140

   1561 GCTGCTGCTGGAGCTGCATTATCTGCTGGGGCTGAAGGTGTAAAACCGGG 1512
        |||||||   :::|||      |||                  |||
    140 ysCysCysCysThrCysThrCysCys.................CysGly 150

   1511 TGCTGCTGCTGCTGCTGCTGGGCCTGCTGC...TGCTGCTGTAGCATGTG 1465
        ||||||||||||:::|||   :::||||||   ||||||||||:::   ||
    151 CysCysCysCysAlaCysAlaCysCysCysAlaCysCysCysThrCysCy 167

   1464 TGTCTCTGGAGTCACT...TTCACTTGGCTAAACAGCACTGCATTGGCAT 1418
        |:::::::||   |||      |||      ::::::|||:::   :::
    167 sThrGlyGlyCysThrCysThrThrCysCysThrGlyThrThrThrThrT 184

   1417 TTGTATGTCCCTGCTGGCTGTGATTCACCTGCTGTTCTAAATTTTTCGTA 1368
          :::|||   |||              |||||||||:::
    184 hrAlaCysThrCys.........CysThrCysCysThrThrThrThrCys 197

   1367 GGTGCTGAAAGAGTCTGTAAGATATGTGCTACATTTGATGGCCGGTTCAT 1318
        ::::::      :::         |||||||:::      |||
    198 AlaThrThrCysAlaThrAlaAlaCysAlaAlaAlaAlaGlyCysThrAl 214

   1317 CTGC 1314
         |||
    214 aCys 215

seq_name: /cqn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-123-155-147

seq_documentation_block:
; Sequence 147, Application US/10123155
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gao,Wei-Qiang
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C30
;  CURRENT APPLICATION NUMBER: US/10/123,155
;  CURRENT FILING DATE:  2002-04-15
; Prior Application removed -  See Palm or File Wrapper
;  NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 147
;   LENGTH: 3734
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-10-123-155-147


alignment_scores:
           Quality:  266.50              Length:    251
             Ratio:  2.562                Gaps:     14
 Percent Similarity:  41.434   Percent Identity: 28.287

alignment_block:
 US-09-664-641-10/rev x US-10-123-155-147  ..

 Align seq 1/1  to: US-10-123-155-147  from: 1  to: 3734

   1920 TGCAGTAACACATCTCTTTCTTTCTCTTATTGCCTGTGCATACGCGCTGC 1871
        |||:::::|||:::               |||  |||      |||::
   2851 CysGlyThrThrAla..............CysCysCysCysCysAlaTh 2862

   1870 TGACTTGACTCTCACAGAGAAGGTGCGTGCATCGACTCGTGAAGGTGGGG 1821
        :      |||                                |||||:
   2862 rGlyCysThr................................GlyGlyT 2868

   1820 TCAACAGTGCCGCCATGTGCCTGGATTATCCTTTTCCAGGTGGCCAGCAG 1771
        ::::::::::|||   |||                  {{{{{ :::::
   2868 hrGlyGlyCysAlaThrCysCysThrCysAlaCysAlaGlyGly.AlaAl 2884

   1770 T............TGCTTATCAGACATCTGCTCTGGATAATCCGCAATTG 1733
        :           |||  :::      |||:::      ::::::  :
   2884 aGlyAlaGlyCysCysThrGlyThrThrCysThrCysCysAlaCysCysT 2901

   1732 CAAACACACATCCCAATAAGAAGCCTTCTTCTGGAATCTCCACTGCTGGA 1683
        :::::                  :::::::|||      |||:::|||
   2901 hrGlyCys............ThrGlyGlyAlaGlyCysCysThrGlyGly 2913

   1682 TCATGTCCAAAAAGCTGATGCTGCTGAGGCGATGGT.............. 1647
        :::|||            ||||||   |||  |||
   2914 AlaCys............CysCysThrGlyGlyGlyGlyThrGlyGlyGl 2926

   1646 ..............GGTGGCTGC.......................... 1638
                      |||||||||
   2926 yAlaCysAlaGlyAlaGlyGlyCysCysThrCysGlyThrCysCysAlaA 2943

   1637 ......................TGTTGCTGCTGC......TGCTGCTGT 1617
                              ||||||||||||     |||||||||
   2943 laCysCysCysCysAlaCysThrCysCysCysCysThrThrCysCysCys 2959

   1616 GAAAATGGA............................TGCGGCGGCTG 1597
        :::|||                              |||   |||||
   2960 GlyThrGlyThrGlyThrCysThrThrCysCysCysCysCysThrGlyCy 2976

   1596 CTGGGGGTAAGGTTGCTGAGAGATCTGCTGCTGCTGC.............. 1560
        |   :::   ||||||         ||||||||||||
   2976 sCysAlaAlaGlyCysCysThrCysCysCysCysCysThrGlyCysCysA 2993

   1559 ........TGCTGCTGGAGCTGCATTATCTGCTGGGGCTGAAGGTGTAAA 1518
                ||||||    |||      |||  |||      |||
   2993 laAlaGlyCysCysThrCysCysCysCysCysThrGlyCysCysCysCys 3009

   1517 ACC..................GGGTGCTGCTGCTGCTGCTGCTGGGCCTG 1486
        |||                  ||||||||||||||||   |||   :::||
   3010 ThrCysThrCysThrGlyAlaGlyCysCysCysCysThrCysGlyCysCy 3026

   1485 CTGCTGCTGCTGTAGCATGTGTGTCTCTGGAGTCACTTTCACTTGGCTAA 1436
        ||||||||||||:::   :::      |||:::      |||      :
   3026 sCysCysCysCysAlaCysAlaCysCysGlyThrCysCysThrCysAlaT 3043
```

```
1435 ACAGCACTGCATTGGCATTTGTATGTCCCTGCTGGCTGTGATTCACCTGC 1386
     ::   |||:::   :::   :::|||   |||             |||
3043 hrCysThrGlyGlyCysCysThrCysCysCys.....CysCys 3055

1385 TGTTCTAAATTTTTCGTAGGTGCTCAAACAGTCTGTAAGATATGTGCTAC 1336
     |||:::           |||:::         |||     |||:::||
3056 CysThr.....GlyGlyCysCysCysCysCysAlaCysThrTh 3068

1335 A 1335
     \
3068 r 3068
```